

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/82, 15/29, 9/00, 5/10, A01H 5/00		A1	(11) International Publication Number: WO 00/09723
			(43) International Publication Date: 24 February 2000 (24.02.00)
(21) International Application Number: PCT/US99/18128 (22) International Filing Date: 10 August 1999 (10.08.99) (30) Priority Data: 60/096,546 14 August 1998 (14.08.98) US (71) Applicant: PIONEER HI-BRED INTERNATIONAL, INC. [US/US]; 800 Capital Square, 400 Locust Street, Des Moines, IA 50309 (US). (72) Inventor: MAHAJAN, Pramod, B.; 8029 Brookview Drive, Urbandale, IA 50322 (US). (74) Agents: YATES, Michael, E. et al.; Darwin Building, 7100 N.W. 62nd Avenue, Johnston, IA 50131-1000 (US).			(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: MAIZE RAD6 GENES AND USES THEREOF			
(57) Abstract The present invention provides methods and compositions relating to altering the Rad6 content of plants. The invention provides isolated nucleic acids and their encoded proteins which are involved in Rad6 biosynthesis. The invention further provides recombinant expression cassettes, host cells, transgenic plants, and antibody compositions.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

MAIZE RAD6 GENES AND USES THEREOF

TECHNICAL FIELD

The present invention relates generally to plant molecular biology. More specifically, it relates to nucleic acids and methods for modulating their expression in plants.

BACKGROUND OF THE INVENTION

Transgenic plant product development by conventional transformation and breeding efforts is a slow and unpredictable process. Gene targeting systems can overcome problems with expression variability, unpredictable of impacts of random gene insertion on agronomic performance, and the large number of experiments that need to be conducted. Such systems can also provide approaches to manipulating endogenous genes. Of course, targeting systems require the ability to focus the recombination process to favor recovery of desired targeting events.

The natural cellular DNA repair and recombination machinery consists of a complex array of protein components interacting in a highly controlled manner to ensure that the fidelity of the genome is conserved throughout the many internal events or external stimuli experienced during each cell cycle. The ability to manipulate this machinery requires an understanding of how specific proteins are involved in the process, and how the genes that encode those proteins are regulated. Since the primary approaches to gene targeting involve recombinases, whether operating in their natural *in vivo* environment (as during normal recombination) or as part of schemes that involve pretreatment of substrates so as to associate DNA with a recombinase and increase efficiency of targeting (e.g., double D-loop), there is a continuing need to isolate and characterize the genes for these molecules. Because many different protein components may be involved in gene targeting, the availability of host-specific genes and proteins could avoid possible problems of incompatibility associated with molecular interactions due to heterologous components. The RAD6 gene of the baker's yeast *Saccharomyces cerevisiae* encodes a protein (designated Rad6) which is involved in DNA repair. The Rad6 protein exhibits ubiquitin conjugating activity. Hence, it is also known as Ubiquitin Conjugating Enzyme (UBC) or E2. The RAD6 gene belongs to a large family which includes many closely related members. RAD6 or E2 has been cloned from many eukaryotic organisms such as worm (*C. elegans*), baker's yeast (*S. cerevisiae*),

budding yeast (*S. pombe*), fruitfly (*D. melanogaster*), and humans. Members of the RAD6 family have also reportedly been cloned from wheat, *Arabidopsis*, tomato, rice, peas and maize. (Stary *et al.*, *Curr. Genet.* 32:309-314, 1997; Gray *et al.*, *Plant Physiol.*, 103:1471-1472, 1993; Feussner *et al.*, *FEBS Lett.* 409:211-215, 1997; Chen and Wang, *Plant Mol. Biol.*, 29: 787-795, 1995; Woo *et al.*, *Gene*, 148:369-370, 1994; Sullivan and Vierstra, *PNAS (USA)* 86: 9861-9865, 1989). The RAD6 gene has been highly conserved in eukaryotic evolution.

RAD6 yeast mutants show very diverse phenotypes indicating the involvement of this gene in multiple cellular functions. Montelone *et al.* showed that a point mutant, *rad6-1*, is proficient in spontaneous and UV induced unequal sister chromatid recombination, but deficient in UV induced mutagenesis, sporulation and meiotic recombination (Montelone BA *et al.*, *Mol. Gen. Genet.*, 184(3):410-415, 1981). These workers also found that another point mutant, *rad6-3*, shared some phenotypes with the *rad6-1* mutant, except that it was not sporulation defective, indicating a specificity of functions. The sporulation defect in *rad6-1* cells appears to be downstream of the premeiotic DNA synthesis but prior to meiotic recombination (Montelone BA *et al.*, *Mol. Gen. Genet.*, 184(3):410-415, 1981). Yeast RAD6 deletion mutants also show a sporulation defect (Game JC *et al.*, *Genetics*, 94: 51-68, 1980). A homologue of RAD6 (termed RPH6⁺) has been identified in the budding yeast *Schizosaccharomyces pombe* (Reynolds P *et al.*, *EMBO J*, 9:1423-30, 1990). The two genes from these distantly related yeast share a high degree of structural and functional homology. Thus, like *rad6*, null mutations of the *rhph6+* gene confer a defect in DNA repair, UV mutagenesis and sporulation, and the RAD6 and *rhph6+* genes can functionally substitute for one another.

Roest *et al.* (Roest *et al.*, *Cell*, 86: 799-810, 1996) generated transgenic mice that were null mutants for RAD6 gene. These mice did not have any detectable levels of RAD6 message or protein. Interestingly and quite unexpectedly, it was observed that the male RAD6 null mice were infertile because of impaired spermatogenesis. More surprisingly, heterozygous males or homozygous females did not show any abnormality with respect to fertility. These findings further substantiated the role of RAD6 genes in inhibiting yeast sporulation.

To date, work with recombinase enzymes in plants, however, has been very limited. Accordingly, there is an ongoing need for the identification and characterization of the functional activities of RAD6 to provide improved and expanded methods of gene targeting in plant systems or for the creation of male sterile plants,

particularly in agriculturally important crop species such as maize. The present invention provides this and other advantages.

SUMMARY OF THE INVENTION

5 Generally, it is the object of the present invention to provide nucleic acids and proteins relating to Rad6. It is an object of the present invention to provide: 1) antigenic fragments of the proteins of the present invention; 2) transgenic plants comprising the nucleic acids of the present invention; 3) methods for modulating, in a transgenic plant, the expression of the nucleic acids of the present invention.

10 Therefore, in one aspect, the present invention relates to an isolated nucleic acid comprising a member selected from the group consisting of (a) a polynucleotide having a specified sequence identity to a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, and 12, wherein the polypeptide when presented as an immunogen elicits the production of an antibody which is specifically
15 reactive to the polypeptide; (b) a polynucleotide which is complementary to the polynucleotide of (a); and (c) a polynucleotide comprising a specific number of contiguous nucleotides from a polynucleotide of (a) or (b). In some embodiments, the polynucleotide has a sequence selected from the group consisting of SEQ ID NOS: 1, 3, 5, 7, 9, and 11. The isolated nucleic acid can be DNA.

20 In another aspect, the present invention relates to recombinant expression cassettes, comprising a nucleic acid as described, *supra*, operably linked to a promoter. In some embodiments, the nucleic acid is operably linked in antisense orientation to the promoter. In another aspect, the present invention is directed to a host cell transfected with the recombinant expression cassette as described, *supra*. In some
25 embodiments, the host cell is a sorghum (*Sorghum bicolor*) or maize (*Zea mays*) cell.

 In a further aspect, the present invention relates to an isolated protein comprising a polypeptide of a specified number of contiguous amino acids encoded by the isolated nucleic acid referred to, *supra*. In some embodiments, the polypeptide has a sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, and 12.

30 In another aspect, the present invention relates to an isolated nucleic acid comprising a polynucleotide of a specified number of nucleotides in length which selectively hybridizes under stringent conditions to a nucleic acid selected from the group consisting of SEQ ID NOS: 1, 3, 5, 7, 9, and 11, or a complement thereof. In some

embodiments, the isolated nucleic acid is operably linked to a promoter.

In yet another aspect, the present invention relates to an isolated nucleic acid comprising a polynucleotide, the polynucleotide having a specified sequence identity to an identical length of a nucleic acid selected from the group consisting of SEQ ID NOS:

5 1, 3, 5, 7, 9, and 11 or a complement thereof.

In another aspect, the present invention relates to an isolated nucleic acid comprising a polynucleotide having a sequence of a nucleic acid amplified from a *Zea mays* nucleic acid library. In some embodiments, the nucleic acid library is a cDNA library.

10 In another aspect, the present invention relates to a recombinant expression cassette comprising a nucleic acid amplified from a library as referred to *supra*, wherein the nucleic acid is operably linked to a promoter. In some embodiments, the present invention relates to a host cell transfected with this recombinant expression cassette. In some embodiments, the present invention relates to a protein of the present invention
15 which is produced from this host cell.

In an additional aspect, the present invention is directed to an isolated nucleic acid comprising a polynucleotide encoding a polypeptide wherein: (a) the polypeptide comprises a specified number of contiguous amino acid residues from a first polypeptide selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, and 12, wherein said
20 polypeptide, when presented as an immunogen, elicits the production of an antibody which specifically binds to said first polypeptide; (b) the polypeptide does not bind to antisera raised against the first polypeptide which has been fully immunosorbed with the first polypeptide; (c) the polypeptide has a molecular weight in non-glycosylated form within a specified percentage of the first polypeptide.

25 In a further aspect, the present invention relates to a heterologous promoter operably linked to a non-isolated polynucleotide of the present invention, wherein the polypeptide is encoded by a nucleic acid amplified from a nucleic acid library.

In yet another aspect, the present invention relates to a transgenic plant comprising a recombinant expression cassette comprising a plant promoter operably
30 linked to any of the isolated nucleic acids of the present invention. In some embodiments, the transgenic plant is *Zea mays*. The present invention also provides transgenic seed from the transgenic plant.

In a further aspect, the present invention relates to a method of modulating

expression of the genes encoding the proteins of the present invention in a plant, comprising the steps of (a) transforming a plant cell with a recombinant expression cassette comprising a polynucleotide of the present invention operably linked to a promoter; (b) growing the plant cell under plant growing conditions; and (c) inducing expression of the polynucleotide for a time sufficient to modulate expression of the genes in the plant. In some embodiments, the plant is maize. Expression of the genes encoding the proteins of the present invention can be increased or decreased relative to a non-transformed control plant.

10 Definitions

Units, prefixes, and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. Numeric ranges are inclusive of the numbers defining the range and include each integer within the defined range. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes. The terms defined below are more fully defined by reference to the specification as a whole.

20 By "amplified" is meant the construction of multiple copies of a nucleic acid sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Cingene, Mississauga, Ontario), Q-Beta
25 Replicase systems, transcription-based amplification system (TAS), and strand displacement amplification (SDA). See, e.g., *Diagnostic Molecular Microbiology: Principles and Applications*, D. H. Persing *et al.*, Ed., American Society for Microbiology, Washington, D.C. (1993). The product of amplification is termed an amplicon.

30 The term "antibody" includes reference to antigen binding forms of antibodies (e.g., Fab, F(ab)₂). The term "antibody" frequently refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). However, while various

antibody fragments can be defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments such as single chain Fv, chimeric antibodies (i.e.,
5 comprising constant and variable regions from different species), humanized antibodies (i.e., comprising a complementarity determining region (CDR) from a non-human source) and heteroconjugate antibodies (e.g., bispecific antibodies).

The term "antigen" includes reference to a substance to which an antibody can be generated and/or to which the antibody is specifically immunoreactive. The specific
10 immunoreactive sites within the antigen are known as epitopes or antigenic determinants. These epitopes can be a linear array of monomers in a polymeric composition - such as amino acids in a protein - or consist of or comprise a more complex secondary or tertiary structure. Those of skill will recognize that all immunogens (i.e., substances capable of eliciting an immune response) are antigens; however some antigens, such as haptens, are
15 not immunogens but may be made immunogenic by coupling to a carrier molecule. An antibody immunologically reactive with a particular antigen can be generated *in vivo* or by recombinant methods such as selection of libraries of recombinant antibodies in phage or similar vectors. See, e.g., Huse *et al.*, *Science* 246: 1275-1281 (1989); and Ward, *et al.*, *Nature* 341: 544-546 (1989); and Vaughan *et al.*, *Nature Biotech.* 14: 309-314
20 (1996).

As used herein, "antisense orientation" includes reference to a duplex polynucleotide sequence which is operably linked to a promoter in an orientation where the antisense strand is transcribed. The antisense strand is sufficiently complementary to an endogenous transcription product such that translation of the endogenous transcription
25 product is often inhibited.

As used herein, "chromosomal region" includes reference to a length of a chromosome which may be measured by reference to the linear segment of DNA which it comprises. The chromosomal region can be defined by reference to two unique DNA sequences, i.e., markers.

30 The term "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or conservatively modified variants of the amino acid sequences. Because of the degeneracy of the genetic

code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations" and represent one species of conservatively modified variation. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of ordinary skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine; and UGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide of the present invention is implicit in each described polypeptide sequence and incorporated herein by reference.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Thus, any number of amino acid residues selected from the group of integers consisting of from 1 to 15 can be so altered. Thus, for example, 1, 2, 3, 4, 5, 7, or 10 alterations can be made. Conservatively modified variants typically provide similar biological activity as the unmodified polypeptide sequence from which they are derived. For example, substrate specificity, enzyme activity, or ligand/receptor binding is generally at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the native protein for its native substrate. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and

6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

See also, Creighton (1984) Proteins W.H. Freeman and Company.

By "encoding" or "encoded", with respect to a specified nucleic acid, is meant comprising the information for translation into the specified protein. A nucleic acid encoding a protein may comprise non-translated sequences (e.g., introns) within translated regions of the nucleic acid, or may lack such intervening non-translated sequences (e.g., as in cDNA). The information by which a protein is encoded is specified by the use of codons. Typically, the amino acid sequence is encoded by the nucleic acid using the "universal" genetic code. However, variants of the universal code, such as are present in some plant, animal, and fungal mitochondria, the bacterium *Mycoplasma capricolum* (*Proc. Natl. Acad. Sci. (USA)*, 82: 2306-2309 (1985)), or the ciliate *Macronucleus*, may be used when the nucleic acid is expressed using these organisms.

When the nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended host where the nucleic acid is to be expressed. For example, although nucleic acid sequences of the present invention may be expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray *et al.* *Nucl. Acids Res.* 17: 477-498 (1989)). Thus, the maize preferred codon for a particular amino acid may be derived from known gene sequences from maize. Maize codon usage for 28 genes from maize plants are listed in Table 4 of Murray *et al.*, *supra*.

As used herein "full-length sequence" in reference to a specified polynucleotide or its encoded protein means having the entire amino acid sequence of, a native (non-synthetic), endogenous, catalytically active form of the specified protein. A full-length sequence can be determined by size comparison relative to a control which is a native (non-synthetic) endogenous cellular form of the specified nucleic acid or protein.

Methods to determine whether a sequence is full-length are well known in the art including such exemplary techniques as northern or western blots, primer extension, S1 protection, and ribonuclease protection. See, e.g., *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997). Comparison to known full-length homologous (orthologous and/or paralogous) sequences can also be used to identify full-length sequences of the present invention. Additionally, consensus

sequences typically present at the 5' and 3' untranslated regions of mRNA aid in the identification of a polynucleotide as full-length. For example, the consensus sequence ANNNNAUGG, where the underlined codon represents the N-terminal methionine, aids in determining whether the polynucleotide has a complete 5' end. Consensus sequences at the 3' end, such as polyadenylation sequences, aid in determining whether the polynucleotide has a complete 3' end.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species or, if from the same species, is substantially modified from its original form by deliberate human intervention.

By "host cell" is meant a cell which contains a vector and supports the replication and/or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells. Preferably, host cells are monocotyledonous or dicotyledonous plant cells. A particularly preferred monocotyledonous host cell is a maize host cell.

The term "hybridization complex" includes reference to a duplex nucleic acid structure formed by two single-stranded nucleic acid sequences selectively hybridized with each other.

By "immunologically reactive conditions" or "immunoreactive conditions" is meant conditions which allow an antibody, generated to a particular epitope, to bind to that epitope to a detectably greater degree (e.g., at least 2-fold over background) than the antibody binds to substantially all other epitopes in a reaction mixture comprising the particular epitope. Immunologically reactive conditions are dependent upon the format of the antibody binding reaction and typically are those utilized in immunoassay protocols. See Harlow and Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York (1988), for a description of immunoassay formats and conditions.

The term "introduced" in the context of inserting a nucleic acid into a cell,

means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

The terms "isolated" refers to material, such as a nucleic acid or a protein, which is: (1) substantially or essentially free from components which normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment; or (2) if the material is in its natural environment, the material has been synthetically (non-naturally) altered by deliberate human intervention to a composition and/or placed at a locus in the cell (e.g., genome or subcellular organelle) not native to a material found in that environment. The alteration to yield the synthetic material can be performed on the material within or removed from its natural state. For example, a naturally occurring nucleic acid becomes an isolated nucleic acid if it is altered, or if it is transcribed from DNA which has been altered, by non-natural, synthetic (i.e., "man-made") methods performed within the cell from which it originates. See, e.g., Compounds and Methods for Site Directed Mutagenesis in Eukaryotic Cells; Kmiec, U.S. Patent No. 5,565,350; *In Vivo* Homologous Sequence Targeting in Eukaryotic Cells; Zarling *et al.*, PCT/US93/03868. Likewise, a naturally occurring nucleic acid (e.g., a promoter) becomes isolated if it is introduced by non-naturally occurring means to a locus of the genome not native to that nucleic acid. Nucleic acids which are "isolated" as defined herein, are also referred to as "heterologous" nucleic acids.

Unless otherwise stated, the term "Rad6 nucleic acid" means a nucleic acid comprising a polynucleotide ("Rad6 polynucleotide") encoding a Rad6 polypeptide. A "Rad6 gene" refers to a non-heterologous genomic form of a full-length Rad6 polynucleotide.

As used herein, "localized within the chromosomal region defined by and including" with respect to particular markers includes reference to a contiguous length of a chromosome delimited by and including the stated markers.

As used herein, "marker" includes reference to a locus on a chromosome that serves to identify a unique position on the chromosome. A "polymorphic marker" includes reference to a marker which appears in multiple forms (alleles) such that

different forms of the marker, when they are present in a homologous pair, allow transmission of each of the chromosomes in that pair to be followed. A genotype may be defined by use of one or a plurality of markers.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids).

By "nucleic acid library" is meant a collection of isolated DNA or RNA molecules which comprise and substantially represent the entire transcribed fraction of a genome of a specified organism. Construction of exemplary nucleic acid libraries, such as genomic and cDNA libraries, is taught in standard molecular biology references such as Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Vol. 152, Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, 2nd ed., Vol. 1-3 (1989); and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, Eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1994 Supplement).

As used herein "operably linked" includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame.

As used herein, the term "plant" includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny of same. Plant cell, as used herein includes, without limitation, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. The class of plants which can be used in the methods of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants. A particularly preferred plant is *Zea mays*.

As used herein, "polynucleotide" includes reference to a

deoxyribopolynucleotide, ribopolynucleotide, or analogs thereof that have the essential nature of a natural ribonucleotide in that they hybridize, under stringent hybridization conditions, to substantially the same nucleotide sequence as naturally occurring nucleotides and/or allow translation into the same amino acid(s) as the naturally occurring nucleotide(s). A polynucleotide can be full-length or a subsequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term includes reference to the specified sequence as well as the complementary sequence thereof. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including *inter alia*, simple and complex cells.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The essential nature of such analogues of naturally occurring amino acids is that, when incorporated into a protein, that protein is specifically reactive to antibodies elicited to the same protein but consisting entirely of naturally occurring amino acids. The terms "polypeptide", "peptide" and "protein" are also inclusive of modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation. Exemplary modifications are described in most basic texts, such as, *Proteins - Structure and Molecular Properties*, 2nd ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, F., *Post-translational Protein Modifications: Perspectives and Prospects*, pp. 1-12 in *Posttranslational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Protein Synthesis: Posttranslational Modifications and Aging*,

Ann. N.Y. Acad. Sci. 663: 48-62 (1992). It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli* or other cells, prior to proteolytic processing, almost invariably will be N-formylmethionine. During post-translational modification of the peptide, a methionine residue at the NH₂-terminus may be deleted. Accordingly, this invention contemplates the use of both the methionine-containing and the methionine-less amino terminal variants of the protein of the invention. In general, as used herein, the term polypeptide encompasses all such modifications, particularly those that are present in polypeptides synthesized by expressing a polynucleotide in a host cell.

As used herein "promoter" includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant promoter" is a promoter capable of initiating transcription in plant cells. Exemplary plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria which comprise genes expressed in plant cells such *Agrobacterium* or *Rhizobium*. Examples of promoters under developmental control include promoters that preferentially initiate transcription in certain tissues, such as leaves, roots, or seeds. Such promoters are referred to as "tissue preferred". Promoters which initiate transcription only in certain tissue are referred to as "tissue specific". A "cell type" specific promoter primarily drives expression in certain cell types in one or more organs, for example, vascular cells in roots or leaves. An "inducible" promoter is a promoter which is under environmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions or the presence of light. Tissue specific, tissue

preferred, cell type specific, and inducible promoters constitute the class of "non-constitutive" promoters. A "constitutive" promoter is a promoter which is active under most environmental conditions.

5 The term "Rad6 polypeptide" refers to one or more amino acid sequences, in glycosylated or non-glycosylated form, encoded by the Rad6 gene. The term is also inclusive of fragments, variants, homologs, alleles or precursors (e.g., preproteins or proproteins) thereof. A "Rad6 protein" comprises a Rad6 polypeptide.

As used herein "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or that the cell is derived
10 from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all as a result of deliberate human intervention. The term "recombinant" as used herein does not encompass the alteration of the cell or vector by naturally occurring
15 events (e.g., spontaneous mutation, natural transformation/transduction/transposition) such as those occurring without deliberate human intervention.

As used herein, a "recombinant expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements which permit transcription of a particular nucleic acid in a host cell. The recombinant
20 expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid to be transcribed, and a promoter.

The term "residue" or "amino acid residue" or "amino acid" are used
25 interchangeably herein to refer to an amino acid that is incorporated into a protein, polypeptide, or peptide (collectively "protein"). The amino acid may be a naturally occurring amino acid and, unless otherwise limited, may encompass known analogs of natural amino acids that can function in a similar manner as naturally occurring amino acids.

30 The term "selectively hybridizes" includes reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of

non-target nucleic acids. Selectively hybridizing sequences typically have about at least 80% sequence identity, preferably 90% sequence identity, and most preferably 100% sequence identity (i.e., complementary) with each other.

The term "specifically reactive", includes reference to a binding reaction between an antibody and a protein having an epitope recognized by the antigen binding site of the antibody. This binding reaction is determinative of the presence of a protein having the recognized epitope amongst the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to an analyte having the recognized epitope to a substantially greater degree (e.g., at least 2-fold over background) than to substantially all other analytes lacking the epitope which are present in the sample.

Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised to the polypeptides of the present invention can be selected from to obtain antibodies specifically reactive with polypeptides of the present invention. The proteins used as immunogens can be in native conformation or denatured so as to provide a linear epitope.

A variety of immunoassay formats may be used to select antibodies specifically reactive with a particular protein (or other analyte). For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York (1988), for a description of immunoassay formats and conditions that can be used to determine selective reactivity.

The terms "stringent conditions" or "stringent hybridization conditions" includes reference to conditions under which a probe will hybridize to its target sequence, to a detectably greater degree than other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length, preferably less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C.

Specificity is typically the function of post-hybridization washes; the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, *Anal. Biochem.*, 138:267-284 (1984): $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ form}) - 500/L$; where M is the molarity of monovalent cations, $\%GC$ is the percentage of guanosine and cytosine nucleotides in the DNA, $\% \text{ form}$ is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. T_m is reduced by about 1 °C for each 1% of mismatching; thus, T_m , hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with $\geq 90\%$ identity are sought, the T_m can be decreased 10 °C. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4 °C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10 °C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20 °C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T_m , those of

ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T_m of less than 45 °C (aqueous solution) or 32 °C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive
5 guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993); and *Current Protocols in Molecular Biology*, Chapter 2, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York
10 (1995).

As used herein, "transgenic plant" includes reference to a plant which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated
15 into the genome alone or as part of a recombinant expression cassette. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term "transgenic" as used herein does not
20 encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

As used herein, "vector" includes reference to a nucleic acid used in transfection
25 of a host cell and into which can be inserted a polynucleotide. Vectors are often replicons. Expression vectors permit transcription of a nucleic acid inserted therein.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", (d) "percentage of sequence identity", and (e)
30 "substantial identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence,

or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" means includes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence may be compared to a reference sequence and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math.* 2: 482 (1981); by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48: 443 (1970); by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci.* 85: 2444 (1988); by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California, GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wisconsin, USA; the CLUSTAL program is well described by Higgins and Sharp, *Gene* 73: 237-244 (1988); Higgins and Sharp, *CABIOS* 5: 151-153 (1989); Corpet, *et al.*, *Nucleic Acids Research* 16: 10881-90 (1988); Huang, *et al.*, *Computer Applications in the Biosciences* 8: 155-65 (1992), and Pearson, *et al.*, *Methods in Molecular Biology* 24: 307-331 (1994). The BLAST family of programs which can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against protein database sequences; BLASTP for protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences; and TBLASTX for nucleotide query sequences against nucleotide database sequences. See, *Current Protocols in Molecular Biology*, Chapter 19, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995).

GAP uses the algorithm of Needleman and Wunsch (J. Mol. Biol. 48: 443-453, 1970) to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty values in Version 10 of the Wisconsin Genetics Software Package for protein sequences are 8 and 2, respectively. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 200. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (see Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters (Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402, 1997; Altschul *et al.*, *J. Mol. Bio.* 215: 403-410, 1990) or to the value obtained using the GAP program using default parameters (see the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wisconsin, USA).

As those of ordinary skill in the art will understand, BLAST searches assume that proteins can be modeled as random sequences. However, many real proteins comprise regions of nonrandom sequences which may be homopolymeric tracts, short-period repeats, or regions enriched in one or more amino acids. Such low-complexity regions may be aligned between unrelated proteins even though other regions of the protein are entirely dissimilar. A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For example, the SEG (Wooten and Federhen, *Comput. Chem.*, 17:149-163 (1993)) and XNU (Claverie and States, *Comput. Chem.*, 17:191-201 (1993)) low-complexity filters can be employed alone or in combination.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences which are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore do not change the functional properties of the molecule.

Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences which differ by such conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well-known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to the algorithm of Meyers and Miller, *Computer Applic. Biol. Sci.*, 4: 11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which

does not comprise additions or deletions) for optimal alignment of the two sequences.

The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total

5 number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

(e) (i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70% sequence identity, preferably at least 80%, more preferably at least 90% and most preferably at least 95%,
10 compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like. Substantial identity of amino acid sequences for these purposes
15 normally means sequence identity of at least 60%, more preferably at least 70%, 80%, 90%, and most preferably at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. However, nucleic acids which do not hybridize to each other under stringent conditions are still substantially
20 identical if the polypeptides which they encode are substantially identical. This may occur, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is that the polypeptide which the first nucleic acid encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

25 (e) (ii) The terms "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70% sequence identity to a reference sequence, preferably 80%, more preferably 85%, most preferably at least 90% or 95% sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of
30 Needleman and Wunsch, *J. Mol. Biol.* 48: 443 (1970). An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative

substitution. Peptides which are "substantially similar" share sequences as noted above except that residue positions which are not identical may differ by conservative amino acid changes.

5 DETAILED DESCRIPTION OF THE INVENTION

Overview

The present invention provides, *inter alia*, compositions and methods for modulating (i.e., increasing or decreasing) the total levels of proteins of the present invention and/or altering their ratios in plants. Thus, the present invention provides
10 utility in such exemplary applications as modulating recombination so as to affect gene targeting or modulating levels of the protein in plant tissue, such as microspores, to decrease male fertility. In particular, the polypeptides of the present invention can be expressed at developmental stages, in tissues, and/or in quantities that are uncharacteristic of non-recombinant plants.

15 The present invention also provides isolated nucleic acid comprising polynucleotides of sufficient length and complementarity to a Rad6 gene to use as probes or amplification primers in the detection, quantitation, or isolation of gene transcripts. For example, isolated nucleic acids of the present invention can be used as probes in detecting deficiencies in the level of mRNA in screenings for desired transgenic plants,
20 for detecting mutations in the gene (e.g., substitutions, deletions, or additions), for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants (polymorphisms) of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of Rad6
25 polypeptides, or for use as immunogens in the preparation and/or screening of antibodies. The isolated nucleic acids of the present invention can also be employed for use in sense or antisense suppression of one or more Rad6 genes in a host cell, tissue, or plant. Attachment of chemical agents which bind, intercalate, cleave and/or crosslink to the isolated nucleic acids of the present invention can also be used to modulate
30 transcription or translation. Further, using a primer specific to an insertion sequence (e.g., transposon) and a primer which specifically hybridizes to an isolated nucleic acid of the present invention, one can use nucleic acid amplification to identify insertion sequence inactivated Rad6 genes from a cDNA library prepared from insertion sequence

mutagenized plants. Progeny seed from the plants comprising the desired inactivated gene can be grown to a plant to study the phenotypic changes characteristic of that inactivation. See, *Tools to Determine the Function of Genes*, 1995 Proceedings of the Fiftieth Annual Corn and Sorghum Industry Research Conference, American Seed Trade Association, Washington, D.C., 1995. Additionally, non-translated 5' or 3' regions of the polynucleotides of the present invention can be used to modulate turnover of heterologous mRNAs and/or protein synthesis. Further, the codon preference characteristic of the polynucleotides of the present invention can be employed in heterologous sequences, or altered in homologous or heterologous sequences, to modulate translational level and/or rates.

The present invention also provides: 1) isolated proteins comprising polypeptides including an amino acid sequence from the Rad6 polypeptides (e.g., preproenzyme, proenzyme, or enzymes) as disclosed herein; 2) proteins comprising at least one epitope from a Rad6 polypeptide. The proteins of the present invention can be employed in assays for enzyme agonists or antagonists of enzyme function, or for use as immunogens or antigens to obtain antibodies specifically immunoreactive with a protein of the present invention. Such antibodies can be used in assays for expression levels, for identifying and/or isolating nucleic acids of the present invention from expression libraries, or for purification of Rad6 polypeptides.

The isolated nucleic acids of the present invention can be used over a broad range of plant types, including species from the genera *Cucurbita*, *Rosa*, *Vitis*, *Juglans*, *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Ciahorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Heterocallis*, *Nemesis*, *Pelargonium*, *Panieum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Pisum*, *Phaseolus*, *Lolium*, *Oryza*, *Zea*, *Avena*, *Hordeum*, *Secale*, *Triticum*, *Sorghum*, *Picea*, and *Populus*.

Nucleic Acids

The present invention provides, *inter alia*, isolated nucleic acids of RNA, DNA, and analogs and/or chimeras thereof, comprising a Rad6 polynucleotide.

The Rad6 nucleic acids of the present invention comprise an isolated Rad6 polynucleotide which is inclusive of:

(a) a polynucleotide encoding a Rad6 polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, and 12 and conservatively modified and polymorphic variants thereof, including exemplary polynucleotides of SEQ ID NOS: 1, 3, 5, 7, 9, and 11; polynucleotide sequences of the invention also include the maize Rad6 polynucleotide sequences as contained in plasmids deposited with American Type Culture Collection (ATCC) and assigned Accession Number PTA-264.

(b) a polynucleotide which is the product of amplification from a *Zea mays* nucleic acid library using primer pairs which amplify polynucleotides having substantial identity to polynucleotides from amongst those having SEQ ID NOS: 1, 3, 5, 7, 9, and 11; or the sequences as contained in the ATCC deposit assigned Accession Number PTA-264.

(c) a polynucleotide which selectively hybridizes to a polynucleotide of (a) or (b);

(d) a polynucleotide having at least 60% sequence identity with polynucleotides of (a), (b), or (c);

(e) a polynucleotide encoding a protein having a specified number of contiguous amino acids from a prototype polypeptide, wherein the protein is specifically recognized by antisera elicited by presentation of the protein and wherein the protein does not detectably immunoreact to antisera which has been fully immunosorbed with the protein;

(f) complementary sequences of polynucleotides of (a), (b), (c), (d), or (e); and

(g) a polynucleotide comprising at least 15 contiguous nucleotides from a polynucleotide of (a), (b), (c), (d), (e), or (f).

The polynucleotides of SEQ ID NOS: 1,3,5,7,9, and 11, are contained in plasmids deposited with American Type Culture Collection (ATCC) on June 24, 1999 and assigned Accession Number PTA-264. American Type Culture Collection is located at 10801 University Blvd., Manassas, VA 20110-2209.

The ATCC deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. The deposit is provided as a convenience to those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. Section 112. The deposited sequences, as well as the polypeptides encoded by the sequences, are incorporated herein by reference and control in the event of any conflict, such as a sequencing error, with the description in this application.

A. Polynucleotides Encoding A Protein of SEQ ID NOS: 2, 4, 6, 8, 10, and 12 or Conservatively Modified or Polymorphic Variants Thereof

As indicated in (a), *supra*, the present invention provides isolated heterologous nucleic acids comprising a Rad6 polynucleotide, wherein the polynucleotide encodes a Rad6 polypeptide, disclosed herein in SEQ ID NOS: 2, 4, 6, 8, 10, and 12, or conservatively modified or polymorphic variants thereof. Those of skill in the art will recognize that the degeneracy of the genetic code allows for a plurality of polynucleotides to encode for the identical amino acid sequence. Such "silent variations" can be used, for example, to selectively hybridize and detect allelic variants of polynucleotides of the present invention. Accordingly, the present invention includes polynucleotides of SEQ ID NOS: 1, 3, 5, 7, 9, and 11, and the sequences as contained in the ATCC deposit assigned Accession Number PTA-264, and silent variations of polynucleotides encoding a polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, and 12. The present invention further provides isolated nucleic acids comprising polynucleotides encoding conservatively modified variants of a polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, and 12. Conservatively modified variants can be used to generate or select antibodies immunoreactive to the non-variant polypeptide. Additionally, the present invention further provides isolated nucleic acids comprising polynucleotides encoding one or more polymorphic (allelic) variants of polypeptides/polynucleotides. Polymorphisms are frequently used to follow segregation of chromosomal regions in, for example, marker assisted selection methods for crop improvement.

B. Polynucleotides Amplified from a Zea mays Nucleic Acid Library

As indicated in (b), *supra*, the present invention provides isolated nucleic acids comprising Rad6 polynucleotides, wherein the polynucleotides are amplified from a *Zea mays* nucleic acid library. *Zea mays* lines B73, PHRE1, A632, BMS-P2#10, W23, and Mo17 are known and publicly available. Other publicly known and available maize lines can be obtained from the Maize Genetics Cooperation (Urbana, IL). The nucleic acid library may be a cDNA library, a genomic library, or a library generally constructed from nuclear transcripts at any stage of intron processing. Generally, a cDNA nucleic acid library will be constructed to comprise a majority of full-length cDNAs. Often, cDNA libraries will be normalized to increase the representation of relatively rare cDNAs. In preferred embodiments, the cDNA library is constructed using a full-length

cDNA synthesis method. Examples of such methods include Oligo-Capping (Maruyama, K. and Sugano, S. *Gene* 138: 171-174, 1994), Biotinylated CAP Trapper (Carninci, P., Kvan, C., *et al.* *Genomics* 37: 327-336, 1996), and CAP Retention Procedure (Edery, E., Chu, L.L., *et al.* *Molecular and Cellular Biology* 15: 3363-3371, 1995). cDNA

5 synthesis is preferably catalyzed at 50-55°C to prevent formation of RNA secondary structure. Examples of reverse transcriptases that are relatively stable at these temperatures are SuperScript II Reverse Transcriptase (Life Technologies, Inc.), AMV Reverse Transcriptase (Boehringer Mannheim) and RetroAmp Reverse Transcriptase (Epicentre). Rapidly growing tissues, or rapidly dividing cells are preferably used as
10 mRNA sources. Table 1, *infra*, provides exemplary tissues, developmental stages, and tissue treatment for library construction.

The present invention also provides subsequences of full-length nucleic acids. Any number of subsequences can be obtained by reference to SEQ ID NOS: 1, 3, 5, 7, 9, and 11, and using primers which selectively amplify, under stringent conditions to:
15 ~~nucleic acid which flank and comprise a polynucleotide of the present invention, or to a~~

site within a polynucleotide of the present invention and a site within the nucleic acid which comprises it. A variety of methods for obtaining 5' and/or 3' ends is well known in the art. See, e.g., RACE (Rapid Amplification of Complementary Ends) as described
20 in Frohman, M. A., in *PCR Protocols: A Guide to Methods and Applications*, M. A. Innis, D. H. Gelfand, J. J. Sninsky, T. J. White, Eds. (Academic Press, Inc., San Diego, 1990), pp. 28-38.); see also, U.S. Pat. No. 5,470,722, and *Current Protocols in Molecular Biology*, Unit 15.6, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); Frohman and Martin, *Techniques* 1:165 (1989). Thus,
25 the present invention provides Rad6 polynucleotides having the sequence of the Rad6 gene, nuclear transcript, cDNA, or complementary sequences and/or subsequences thereof.

Primer sequences can be obtained by reference to a contiguous subsequence of a polynucleotide of the present invention. Primers are chosen to selectively hybridize,
30 under stringent conditions, to a polynucleotide of the present invention in an amplification mixture comprising a genomic and/or cDNA library from the same species. Generally, the primers are complementary to a subsequence of the amplicon they yield.

In some embodiments, the primers will be constructed to selectively hybridize under stringent conditions at their 5' terminal end's to the codon encoding the carboxy or amino terminal amino acid residue (or the complements thereof) of the polynucleotides of the present invention. In other embodiments the primers will be chosen to selectively hybridize under stringent conditions to a locus within the polynucleotide which comprises the codon encoding the amino terminus (i.e., the 5' terminal coding region) and the codon encoding the carboxy terminus (i.e., the 3' terminal coding region). The primer length in nucleotides is selected from the group of integers consisting of from at least 15 to 50. Thus, the primers can be at least 15, 18, 20, 25, 30, 40, or 50 nucleotides in length. A non-annealing sequence at the 5' end of a primer (a "tail") can be added, for example, to introduce a cloning site at the terminal ends of the amplicon.

The amplification primers may optionally be elongated in the 3' direction with additional contiguous nucleotides from the polynucleotide sequences, such as SEQ ID NOS: 1, 3, 5, 7, 9, and 11, from which they are derived. The number of nucleotides by which the primers can be elongated is selected from the group of integers consisting of from at least 1 to 25. Thus, for example, the primers can be elongated with an additional 1, 5, 10, or 15 nucleotides. Those of skill will recognize that a lengthened primer sequence can be employed to increase specificity of binding (i.e., annealing) to a target sequence.

The amplification products can be translated using expression systems well known to those of skill in the art and as discussed, *infra*. The resulting translation products can be confirmed as polypeptides of the present invention by, for example, assaying for the appropriate catalytic activity (e.g., specific activity and/or substrate specificity), or verifying the presence of one or more linear epitopes which are specific to a polypeptide of the present invention. Methods for protein synthesis from PCR derived templates are known in the art and commercially available. See, e.g., Amersham Life Sciences, Inc, Catalog '97, p.354.

C. Polynucleotides Which Selectively Hybridize to a Polynucleotide of (A) or (B)

As indicated in (c), *supra*, the present invention provides isolated nucleic acids comprising Rad6 polynucleotides, wherein the polynucleotides selectively hybridize, under stringent hybridization conditions, to a polynucleotide of paragraphs (A) or (B) as discussed, *supra*. Thus, the polynucleotides of this embodiment can be used for

isolating, detecting, and/or quantifying nucleic acids comprising the polynucleotides of (A) or (B). For example, polynucleotides of the present invention can be used to identify, isolate, or amplify partial or full-length clones in a deposited library. In some embodiments, the polynucleotides are genomic or cDNA sequences isolated, or otherwise complementary to, a cDNA from a dicot or monocot nucleic acid library. In optional embodiments, the nucleic acid library is constructed from a species of the family *Gramineae* such as *Zea mays*. Exemplary species of monocots and dicots include, but are not limited to: corn, canola, soybean, cotton, wheat, sorghum, sunflower, oats, sugar cane, millet, barley, and rice. Preferably, the cDNA library comprises at least 80% full-length sequences, preferably at least 85% or 90% full-length sequences, and more preferably at least 95% full-length sequences. The cDNA libraries can be normalized to increase the representation of rare sequences. Low stringency hybridization conditions are typically, but not exclusively, employed with sequences having a reduced sequence identity relative to complementary sequences. Moderate and high stringency conditions can optionally be employed for sequences of greater identity. Low stringency conditions allow selective hybridization of sequences having about 70% sequence identity and can be employed to identify orthologous or paralogous sequences.

D. Polynucleotides Having at Least 60% Sequence Identity with the Polynucleotides of (A), (B) or (C)

As indicated in (d), *supra*, the present invention provides isolated nucleic acids comprising Rad6 polynucleotides, wherein the polynucleotides have a specified identity at the nucleotide level to a polynucleotide as disclosed above in paragraphs (A), (B), or (C). The percentage of identity to a reference sequence is at least 60% and, rounded upwards to the nearest integer, can be expressed as an integer selected from the group of integers consisting of from 60 to 99. Thus, for example, the percentage of identity to a reference sequence can be at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%.

Optionally, the polynucleotides of this embodiment will share an epitope with a polypeptide encoded by the polynucleotides of (A), (B), or (C). Thus, these polynucleotides encode a first polypeptide which elicits production of antisera comprising antibodies which are specifically reactive to a second polypeptide encoded by a polynucleotide of (A), (B), or (C). However, the first polypeptide does not bind to

antisera raised against itself when the antisera has been fully immunosorbed with the first polypeptide. Hence, the polynucleotides of this embodiment can be used to generate antibodies for use in, for example, the screening of expression libraries for nucleic acids comprising polynucleotides of (A), (B), or (C), or for purification of, or in
5 immunoassays for, polypeptides encoded by the polynucleotides of (A), (B), or (C). The polynucleotides of this embodiment embrace nucleic acid sequences which can be employed for selective hybridization to a polynucleotide encoding a polypeptide of the present invention.

Screening polypeptides for specific binding to antisera can be conveniently
10 achieved using peptide display libraries. This method involves the screening of large collections of peptides for individual members having the desired function or structure. Antibody screening of peptide display libraries is well known in the art. The displayed peptide sequences can be from 3 to 5000 or more amino acids in length, frequently from 5-100 amino acids long, and often from about 8 to 15 amino acids long. In addition to
15 direct chemical synthetic methods for generating peptide libraries, several recombinant DNA methods have been described. One type involves the display of a peptide sequence on the surface of a bacteriophage or cell. Each bacteriophage or cell contains the nucleotide sequence encoding the particular displayed peptide sequence. Such methods are described in PCT patent publication Nos. 91/17271, 91/18980, 91/19818,
20 and 93/08278. Other systems for generating libraries of peptides have aspects of both *in vitro* chemical synthesis and recombinant methods. See, PCT Patent publication Nos. 92/05258, 92/14843, and 96/19256. See also, U.S. Patent Nos. 5,658,754; and 5,643,768. Peptide display libraries, vectors, and screening kits are commercially available from such suppliers as Invitrogen (Carlsbad, CA).

25

E. Polynucleotides Encoding a Protein Having a Subsequence from a Prototype Polypeptide and is Cross-Reactive to the Prototype Polypeptide

As indicated in (e), *supra*, the present invention provides isolated nucleic acids comprising Rad6 polynucleotides, wherein the polynucleotides encode a protein having a
30 subsequence of contiguous amino acids from a prototype Rad6 polypeptide. Exemplary prototype Rad6 polypeptides are provided in SEQ ID NOS: 2, 4, 6, 8, 10, and 12. The length of contiguous amino acids from the prototype polypeptide is selected from the group of integers consisting of from at least 10 to the number of amino acids within the

prototype sequence. Thus, for example, the polynucleotide can encode a polypeptide having a subsequence having at least 10, 15, 20, 25, 30, 35, 40, 45, or 50, contiguous amino acids from the prototype polypeptide. Further, the number of such subsequences encoded by a polynucleotide of the instant embodiment can be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5. The subsequences can be separated by any integer of nucleotides from 1 to the number of nucleotides in the sequence such as at least 5, 10, 15, 25, 50, 100, or 200 nucleotides.

The proteins encoded by polynucleotides of this embodiment, when presented as an immunogen, elicit the production of polyclonal antibodies which specifically bind to a prototype polypeptide such as, but not limited to, a polypeptide encoded by the polynucleotide of (b), *supra*, or exemplary polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, and 12. Generally, however, a protein encoded by a polynucleotide of this embodiment does not bind to antisera raised against the prototype polypeptide when the antisera has been fully immunosorbed with the prototype polypeptide. Methods of making and assaying for antibody binding specificity/affinity are well known in the art. Exemplary immunoassay formats include ELISA, competitive immunoassays, radioimmunoassays, Western blots, indirect immunofluorescent assays and the like.

In a preferred assay method, fully immunosorbed and pooled antisera which is elicited to the prototype polypeptide can be used in a competitive binding assay to test the protein. The concentration of the prototype polypeptide required to inhibit 50% of the binding of the antisera to the prototype polypeptide is determined. If the amount of the protein required to inhibit binding is less than twice the amount of the prototype protein, then the protein is said to specifically bind to the antisera elicited to the immunogen. Accordingly, the proteins of the present invention embrace allelic variants, conservatively modified variants, and minor recombinant modifications to a prototype polypeptide.

A polynucleotide of the present invention optionally encodes a protein having a molecular weight as the non-glycosylated protein within 20% of the molecular weight of the full-length non-glycosylated Rad6 polypeptides as disclosed herein (e.g., SEQ ID NOS: 2, 4, 6, 8, 10, and 12). Molecular weight can be readily determined by SDS-PAGE under reducing conditions. Preferably, the molecular weight is within 15% of a full length Rad6 polypeptide, more preferably within 10% or 5%, and most preferably within 3%, 2%, or 1% of a full length Rad6 polypeptide of the present invention.

Molecular weight determination of a protein can be conveniently performed by SDS-PAGE under denaturing conditions.

Optionally, the polynucleotides of this embodiment will encode a protein having a specific activity at least 20%, 30%, 40%, or 50% of the native, endogenous (i.e., non-isolated), full-length Rad6 polypeptide. Further, the proteins encoded by polynucleotides of this embodiment will optionally have a substantially similar affinity constant (K_m) and/or catalytic activity (i.e., the microscopic rate constant, k_{cat}) as the native endogenous, full-length Rad6 protein. Those of skill in the art will recognize that k_{cat}/K_m value determines the specificity for competing substrates and is often referred to as the specificity constant. Proteins of this embodiment can have a k_{cat}/K_m value at least 10% of the non-isolated full-length Rad6 polypeptide as determined using the substrate of that polypeptide from the Rad6 specific pathways, *supra*. Optionally, the k_{cat}/K_m value will be at least 20%, 30%, 40%, 50%, and most preferably at least 60%, 70%, 80%, 90%, or 95% the k_{cat}/K_m value of the non-isolated, full-length Rad6 polypeptide.

Determination of k_{cat} , K_m , and k_{cat}/K_m can be determined by any number of means well known to those of skill in the art. For example, the initial rates (i.e., the first 5% or less of the reaction) can be determined using rapid mixing and sampling techniques (e.g., continuous-flow, stopped-flow, or rapid quenching techniques), flash photolysis, or relaxation methods (e.g., temperature jumps) in conjunction with such exemplary methods of measuring as spectrophotometry, spectrofluorimetry, nuclear magnetic resonance, or radioactive procedures. Kinetic values are conveniently obtained using a Lineweaver-Burk or Eadie-Hofstee plot.

F. Polynucleotides Complementary to the Polynucleotides of (A)-(E)

As indicated in (f), *supra*, the present invention provides isolated nucleic acids comprising Rad6 polynucleotides, wherein the polynucleotides are complementary to the polynucleotides of paragraphs A-E, above. As those of skill in the art will recognize, complementary sequences base-pair throughout the entirety of their length with the polynucleotides of (A)-(E) (i.e., have 100% sequence identity over their entire length).

Complementary bases associate through hydrogen bonding in double stranded nucleic acids. For example, the following base pairs are complementary: guanine and cytosine; adenine and thymine; and adenine and uracil.

G. Polynucleotides Which are Subsequences of the Polynucleotides of (A)-(F)

As indicated in (g), *supra*, the present invention provides isolated nucleic acids comprising Rad6 polynucleotides, wherein the polynucleotide comprises at least 15 contiguous bases from the polynucleotides of (A) through (F) as discussed above. The length of the polynucleotide is given as an integer selected from the group consisting of from at least 15 to the length of the nucleic acid sequence from which the polynucleotide is a subsequence of. Thus, for example, polynucleotides of the present invention are inclusive of polynucleotides comprising at least 15, 20, 25, 30, 40, 50, 60, 75, or 100 contiguous nucleotides in length from the polynucleotides of (A)-(F). Optionally, the number of such subsequences encoded by a polynucleotide of the instant embodiment can be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5. The subsequences can be separated by any integer of nucleotides from 1 to the number of nucleotides in the sequence such as at least 5, 10, 15, 25, 50, 100, or 200 nucleotides.

The subsequences of the present invention can comprise structural characteristics of the sequence from which it is derived. Alternatively, the subsequences can lack certain structural characteristics of the larger sequence from which it is derived. For example, a subsequence from a polynucleotide encoding a polypeptide having at least

one linear epitope in common with a prototype sequence, such as SEQ ID NOS: 2, 4, 6, 8, 10, and 12, may encode an epitope in common with the prototype sequence.

Alternatively, the subsequence may not encode an epitope in common with the prototype sequence but can be used to isolate the larger sequence by, for example, nucleic acid hybridization with the sequence from which it's derived. Subsequences can be used to modulate or detect gene expression by introducing into the subsequences compounds which bind, intercalate, cleave and/or crosslink to nucleic acids. Exemplary compounds include acridine, psoralen, phenanthroline, naphthoquinone, daunomycin or chloroethylaminoaryl conjugates.

Construction of Nucleic Acids

The isolated nucleic acids of the present invention can be made using (a) standard recombinant methods, (b) synthetic techniques, or combinations thereof. In some embodiments, the polynucleotides of the present invention will be cloned, amplified, or otherwise constructed from a monocot. In preferred embodiments the monocot is *Zea*

mays. Particularly preferred is the use of *Zea mays* tissue, preferably those disclosed in Table 1, *infra*.

The nucleic acids may conveniently comprise sequences in addition to a polynucleotide of the present invention. For example, a multi-cloning site comprising one or more endonuclease restriction sites may be inserted into the nucleic acid to aid in isolation of the polynucleotide. Also, translatable sequences may be inserted to aid in the isolation of the translated polynucleotide of the present invention. For example, a hexa-histidine marker sequence provides a convenient means to purify the proteins of the present invention. The nucleic acid of the present invention - excluding the polynucleotide sequence - is optionally a vector, adapter, or linker for cloning and/or expression of a polynucleotide of the present invention. Additional sequences may be added to such cloning and/or expression sequences to optimize their function in cloning and/or expression, to aid in isolation of the polynucleotide, or to improve the introduction of the polynucleotide into a cell. Typically, the length of a nucleic acid of the present invention less the length of its polynucleotide of the present invention is less than 20 kilobase pairs, often less than 15 kb, and frequently less than 10 kb. Use of cloning vectors, expression vectors, adapters, and linkers is well known in the art. Exemplary nucleic acids include such vectors as: M13, lambda ZAP Express, lambda ZAP II, lambda gt10, lambda gt11, pBK-CMV, pBK-RSV, pBluescript II, lambda DASH II, lambda EMBL 3, lambda EMBL 4, pWE15, SuperCos 1, SurfZap, Uni-ZAP, pBC, pBS+/-, pSG5, pBK, pCR-Script, pET, pSPUTK, p3'SS, pOPRSVI CAT, pOPI3 CAT, pXT1, pSG5, pPbac, pMbac, pMC1neo, pOG44, pOG45, pFRT β GAL, pNEO β GAL, pRS403, pRS404, pRS405, pRS406, pRS413, pRS414, pRS415, pRS416, lambda MOSSlox, and lambda MOSElox. For a description of various nucleic acids see, for example, Stratagene Cloning Systems, Catalogs 1995, 1996, 1997 (La Jolla, CA); and, Amersham Life Sciences, Inc, Catalog '97 (Arlington Heights, IL).

A. Recombinant Methods for Constructing Nucleic Acids

The isolated nucleic acid compositions of this invention, such as RNA, cDNA, genomic DNA, or a hybrid thereof, can be obtained from plant biological sources using any number of cloning methodologies known to those of skill in the art. In some embodiments, oligonucleotide probes which selectively hybridize, under stringent conditions, to the polynucleotides of the present invention are used to identify the desired

sequence in a cDNA or genomic DNA library. While isolation of RNA, and construction of cDNA and genomic libraries is well known to those of ordinary skill in the art, the following highlights some of the methods employed.

5

A1. mRNA Isolation and Purification

Total RNA from plant cells comprises such nucleic acids as mitochondrial RNA, chloroplastic RNA, rRNA, tRNA, hnRNA and mRNA. Total RNA preparation typically involves lysis of cells and removal of proteins, followed by precipitation of nucleic acids. Extraction of total RNA from plant cells can be accomplished by a variety of means. Frequently, extraction buffers include a strong detergent such as SDS and an organic denaturant such as guanidinium isothiocyanate, guanidine hydrochloride or phenol. Following total RNA isolation, poly(A)⁺ mRNA is typically purified from the remainder RNA using oligo(dT) cellulose. Exemplary total RNA and mRNA isolation protocols are described in *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); and, *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995). Total RNA and mRNA isolation kits are commercially available from vendors such as Stratagene (La Jolla, CA), Clontech (Palo Alto, CA), Pharmacia (Piscataway, NJ), and 5'-3' (Paoli, PA). See also, U.S. Patent Nos. 5,614,391; and, 5,459,253. The mRNA can be fractionated into populations with size ranges of about 0.5, 1.0, 1.5, 2.0, 2.5 or 3.0 kb. The cDNA synthesized for each of these fractions can be size selected to the same size range as its mRNA prior to vector insertion. This method helps eliminate truncated cDNA formed by incompletely reverse transcribed mRNA.

25

A2. Construction of a cDNA Library

Construction of a cDNA library generally entails five steps. First, first strand cDNA synthesis is initiated from a poly(A)⁺ mRNA template using a poly(dT) primer or random hexanucleotides. Second, the resultant RNA-DNA hybrid is converted into double stranded cDNA, typically by a combination of RNase H and DNA polymerase I (or Klenow fragment). Third, the termini of the double stranded cDNA are ligated to adaptors. Ligation of the adaptors will produce cohesive ends for cloning. Fourth, size selection of the double stranded cDNA eliminates excess adaptors and primer fragments,

and eliminates partial cDNA molecules due to degradation of mRNAs or the failure of reverse transcriptase to synthesize complete first strands. Fifth, the cDNAs are ligated into cloning vectors and packaged. cDNA synthesis protocols are well known to the skilled artisan and are described in such standard references as: *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); and, *Current Protocols in Molecular Biology*, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995). cDNA synthesis kits are available from a variety of commercial vendors such as Stratagene or Pharmacia.

10 A number of cDNA synthesis protocols have been described which provide substantially pure full-length cDNA libraries. Substantially pure full-length cDNA libraries are constructed to comprise at least 90%, and more preferably at least 93% or 95% full-length inserts amongst clones containing inserts. The length of insert in such libraries can be from 0 to 8, 9, 10, 11, 12, 13, or more kilobase pairs. Vectors to accommodate inserts of these sizes are known in the art and available commercially. See, 15 e.g., Stratagene's lambda ZAP Express (cDNA cloning vector with 0 to 12 kb cloning capacity).

An exemplary method of constructing a greater than 95% pure full-length cDNA library is described by Carninci et al., *Genomics*, 37:327-336 (1996). In that protocol, the cap-structure of eukaryotic mRNA is chemically labeled with biotin. By using 20 streptavidin-coated magnetic beads, only the full-length first-strand cDNA/mRNA hybrids are selectively recovered after RNase I treatment. The method provides a high yield library with an unbiased representation of the starting mRNA population. Other methods for producing full-length libraries are known in the art. See, e.g., Edery et al., *Mol. Cell Biol.*, 15(6):3363-3371 (1995); and, PCT Application WO 96/34981.

25

A3. Normalized or Subtracted cDNA Libraries

A non-normalized cDNA library represents the mRNA population of the tissue it was made from. Since unique clones are out-numbered by clones derived from highly expressed genes their isolation can be laborious. Normalization of a cDNA library is the 30 process of creating a library in which each clone is more equally represented.

A number of approaches to normalize cDNA libraries are known in the art. One approach is based on hybridization to genomic DNA. The frequency of each hybridized cDNA in the resulting normalized library would be proportional to that of each

corresponding gene in the genomic DNA. Another approach is based on kinetics. If cDNA reannealing follows second-order kinetics, rarer species anneal less rapidly and the remaining single-stranded fraction of cDNA becomes progressively more normalized during the course of the hybridization. Specific loss of any species of cDNA, regardless of its abundance, does not occur at any Cot value. Construction of normalized libraries is described in Ko, *Nucl. Acids. Res.*, 18(19):5705-5711 (1990); Patanjali *et al.*, *Proc. Natl. Acad. U.S.A.*, 88:1943-1947 (1991); U.S. Patents 5,482,685, and 5,637,685. In an exemplary method described by Soares *et al.*, normalization resulted in reduction of the abundance of clones from a range of four orders of magnitude to a narrow range of only 1 order of magnitude. *Proc. Natl. Acad. Sci. USA*, 91:9228-9232 (1994).

Subtracted cDNA libraries are another means to increase the proportion of less abundant cDNA species. In this procedure, cDNA prepared from one pool of mRNA is depleted of sequences present in a second pool of mRNA by hybridization. The cDNA:mRNA hybrids are removed and the remaining un-hybridized cDNA pool is enriched for sequences unique to that pool. See, Foote *et al.* in, *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); Kho and Zarbl, *Technique*, 3(2):58-63 (1991); Sive and St. John, *Nucl. Acids Res.*, 16(22):10937 (1988); *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); and, Swaroop *et al.*, *Nucl. Acids Res.*, 19(8):1954 (1991). cDNA subtraction kits are commercially available. See, e.g., PCR-Select (Clontech).

A4. Construction of a Genomic Library

To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. Methodologies to accomplish these ends, and sequencing methods to verify the sequence of nucleic acids are well known in the art. Examples of appropriate molecular biological techniques and instructions sufficient to direct persons of skill through many construction, cloning, and screening methodologies are found in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Vols. 1-3 (1989), *Methods in Enzymology*, Vol. 152: *Guide to Molecular Cloning Techniques*, Berger and Kimmel, Eds., San Diego: Academic Press, Inc. (1987), *Current*

Protocols in Molecular Biology, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997). Kits for construction of genomic libraries are also commercially available.

5

A5. Nucleic Acid Screening and Isolation Methods

The cDNA or genomic library can be screened using a probe based upon the sequence of a polynucleotide of the present invention such as those disclosed herein. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Those of skill in the art will appreciate that various degrees of stringency of hybridization can be employed in the assay; and either the hybridization or the wash medium can be stringent. As the conditions for hybridization become more stringent, there must be a greater degree of complementarity between the probe and the target for duplex formation to occur. The degree of stringency can be controlled by temperature, ionic strength, pH and the presence of a partially denaturing solvent such as formamide. For example, the stringency of hybridization is conveniently varied by changing the polarity of the reactant solution through manipulation of the concentration of formamide within the range of 0% to 50%. The degree of complementarity (sequence identity) required for detectable binding will vary in accordance with the stringency of the hybridization medium and/or wash medium. The degree of complementarity will optimally be 100 percent; however, it should be understood that minor sequence variations in the probes and primers may be compensated for by reducing the stringency of the hybridization and/or wash medium.

The nucleic acids of interest can also be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of polynucleotides of the present invention and related genes directly from genomic DNA or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, U.S. Patent No. 4,683,202 (1987); and, PCR

Protocols A Guide to Methods and Applications, Innis *et al.*, Eds., Academic Press Inc., San Diego, CA (1990). Commercially available kits for genomic PCR amplification are known in the art. See, e.g., Advantage-GC Genomic PCR Kit (Clontech). The T4 gene 32 protein (Boehringer Mannheim) can be used to improve yield of long PCR products.

5 PCR-based screening methods have also been described. Wilfinger *et al.* describe a PCR-based method in which the longest cDNA is identified in the first step so that incomplete clones can be eliminated from study. *BioTechniques*, 22(3): 481-486 (1997). In that method, a primer pair is synthesized with one primer annealing to the 5' end of the sense strand of the desired cDNA and the other primer to the vector. Clones
10 are pooled to allow large-scale screening. By this procedure, the longest possible clone is identified amongst candidate clones. Further, the PCR product is used solely as a diagnostic for the presence of the desired cDNA and does not utilize the PCR product itself. Such methods are particularly effective in combination with a full-length cDNA construction methodology, *supra*.

15 B. Synthetic Methods for Constructing Nucleic Acids

The isolated nucleic acids of the present invention can also be prepared by direct chemical synthesis by methods such as the phosphotriester method of Narang *et al.*, *Meth. Enzymol.* 68: 90-99 (1979); the phosphodiester method of Brown *et al.*, *Meth.*
20 *Enzymol.* 68: 109-151 (1979); the diethylphosphoramidite method of Beaucage *et al.*, *Tetra. Lett.* 22: 1859-1862 (1981); the solid phase phosphoramidite triester method described by Beaucage and Caruthers, *Tetra. Letts.* 22(20): 1859-1862 (1981), *e.g.*, using an automated synthesizer, *e.g.*, as described in Needham-VanDevanter *et al.*, *Nucleic Acids Res.*, 12: 6159-6168 (1984); and, the solid support method of U.S. Patent
25 No. 4,458,066. Chemical synthesis generally produces a single stranded oligonucleotide. This may be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill will recognize that while chemical synthesis of DNA is limited to sequences of about 100 bases, longer sequences may be obtained by
30 the ligation of shorter sequences.

Recombinant Expression Cassettes

The present invention further provides recombinant expression cassettes comprising a nucleic acid of the present invention. A nucleic acid sequence coding for the desired polynucleotide of the present invention, for example a cDNA or a genomic sequence encoding a full length polypeptide of the present invention, can be used to
5 construct a recombinant expression cassette which can be introduced into the desired host cell. A recombinant expression cassette will typically comprise a polynucleotide of the present invention operably linked to transcriptional initiation regulatory sequences which will direct the transcription of the polynucleotide in the intended host cell, such as tissues of a transformed plant.

10 For example, plant expression vectors may include (1) a cloned plant gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific/selective expression), a
15 transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

A plant promoter fragment can be employed which will direct expression of a polynucleotide of the present invention in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most
20 environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the ubiquitin 1 promoter, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Patent No. 5,683,439), the *Nos* promoter, the pEmu
25 promoter, the rubisco promoter, the GRP1-8 promoter, and other transcription initiation regions from various plant genes known to those of skill.

Alternatively, the plant promoter can direct expression of a polynucleotide of the present invention in a specific tissue or may be otherwise under more precise environmental or developmental control. Such promoters are referred to here as
30 "inducible" promoters. Environmental conditions that may effect transcription by inducible promoters include pathogen attack, anaerobic conditions, or the presence of light. Examples of inducible promoters are the *Adh1* promoter which is inducible by

hypoxia or cold stress, the Hsp70 promoter which is inducible by heat stress, and the PPKK promoter which is inducible by light.

Examples of promoters under developmental control include promoters that initiate transcription only, or preferentially, in certain tissues, such as leaves, roots, fruit, seeds, or flowers. An exemplary promoter is the anther specific promoter 5126 (U.S. Patent Nos. 5,689,049 and 5,689,051). The operation of a promoter may also vary depending on its location in the genome. Thus, an inducible promoter may become fully or partially constitutive in certain locations.

Both heterologous and non-heterologous (i.e., endogenous) promoters can be employed to direct expression of the nucleic acids of the present invention. These promoters can also be used, for example, in recombinant expression cassettes to drive expression of antisense nucleic acids to reduce, increase, or alter Rad6 content and/or composition in a desired tissue, or to generate male sterile plants. Rad6 or E2 nucleic acids from a variety of sources, as discussed *supra*, can be employed to create male sterile plants. Exemplary Rad6 nucleic acids may be derived or isolated from yeast or mouse to create male sterile plants. In optional embodiments the Rad6 or E2 gene or cDNA is operably linked to an anther specific promoter such as 5126, *supra*.

Preferably, the male sterile plant is maize.

Thus, in some embodiments, the nucleic acid construct will comprise a promoter functional in a plant cell, such as in *Zea mays*, operably linked to a polynucleotide of the present invention. Promoters useful in these embodiments include the endogenous promoters driving expression of a polypeptide of the present invention.

In some embodiments, isolated nucleic acids which serve as promoter or enhancer elements can be introduced in the appropriate position (generally upstream) of a non-heterologous form of a polynucleotide of the present invention so as to up or down regulate expression of a polynucleotide of the present invention. For example, endogenous promoters can be altered *in vivo* by mutation, deletion, and/or substitution (see, Kmiec, U.S. Patent 5,565,350; Zarling *et al.*, PCT/US93/03868), or isolated promoters can be introduced into a plant cell in the proper orientation and distance from a Rad6 gene so as to control the expression of the gene. Gene expression can be modulated under conditions suitable for plant growth so as to alter Rad6 content and/or composition. Thus, the present invention provides compositions, and methods for

making, heterologous promoters and/or enhancers operably linked to a native, endogenous (i.e., non-heterologous) form of a polynucleotide of the present invention.

Methods for identifying promoters with a particular expression pattern, in terms of, e.g., tissue type, cell type, stage of development, and/or environmental conditions, are well known in the art. See, e.g., *The Maize Handbook*, Chapters 114-115; Freeling and Walbot, Eds., Springer, New York (1994); *Corn and Corn Improvement*, 3rd edition, Chapter 6, Sprague and Dudley, Eds., American Society of Agronomy, Madison, Wisconsin (1988). A typical step in promoter isolation methods is identification of gene products that are expressed with some degree of specificity in the target tissue. Amongst the range of methodologies are: differential hybridization to cDNA libraries; subtractive hybridization; differential display; differential 2-D protein gel electrophoresis; DNA probe arrays; and isolation of proteins known to be expressed with some specificity in the target tissue. Such methods are well known to those of skill in the art.

Commercially available products for identifying promoters are known in the art such as Clontech's (Palo Alto, CA) Universal GenomeWalker Kit.

For the protein-based methods, it is helpful to obtain the amino acid sequence for at least a portion of the identified protein, and then to use the protein sequence as the basis for preparing a nucleic acid that can be used as a probe to identify either genomic DNA directly, or preferably, to identify a cDNA clone from a library prepared from the target tissue. Once such a cDNA clone has been identified, that sequence can be used to identify the sequence at the 5' end of the transcript of the indicated gene. For differential hybridization, subtractive hybridization and differential display, the nucleic acid sequence identified as enriched in the target tissue is used to identify the sequence at the 5' end of the transcript of the indicated gene. Once such sequences are identified, starting either from protein sequences or nucleic acid sequences, any of these sequences identified as being from the gene transcript can be used to screen a genomic library prepared from the target organism. Methods for identifying and confirming the transcriptional start site are well known in the art.

In the process of isolating promoters expressed under particular environmental conditions or stresses, or in specific tissues, or at particular developmental stages, a number of genes are identified that are expressed under the desired circumstances, in the desired tissue, or at the desired stage. Further analysis will reveal expression of each particular gene in one or more other tissues of the plant. One can identify a promoter

with activity in the desired tissue or condition but that do not have activity in any other common tissue.

To identify the promoter sequence, the 5' portions of the clones described here are analyzed for sequences characteristic of promoter sequences. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually an AT-rich stretch of 5-10 bp located approximately 20 to 40 base pairs upstream of the transcription start site. Identification of the TATA box is well known in the art. For example, one way to predict the location of this element is to identify the transcription start site using standard RNA-mapping techniques such as primer extension, S1 analysis, and/or RNase protection. To confirm the presence of the AT-rich sequence, a structure-function analysis can be performed involving mutagenesis of the putative region and quantification of the mutation's effect on expression of a linked downstream reporter gene. See, e.g., *The Maize Handbook*, Chapter 114, Freeling and Walbot, Eds., Springer, New York, (1994).

In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element (i.e., the CAAT box) with a series of adenines surrounding the trinucleotide G (or T) N G. J. Messing *et al.*, in *Genetic Engineering in Plants*, Kosage, Meredith and Hollaender, Eds., pp. 221-227, 1983. In maize, there is no well-conserved CAAT box but there are several short, conserved protein-binding motifs upstream of the TATA box. These include motifs for the trans-acting transcription factors involved in light regulation, anaerobic induction, hormonal regulation, or anthocyanin biosynthesis, as appropriate for each gene.

Once promoter and/or gene sequences are known, a region of suitable size is selected from the genomic DNA that is 5' to the transcriptional start, or the translational start site, and such sequences are then linked to a coding sequence. If the transcriptional start site is used as the point of fusion, any of a number of possible 5' untranslated regions can be used in between the transcriptional start site and the partial coding sequence. If the translational start site at the 3' end of the specific promoter is used, then it is linked directly to the methionine start codon of a coding sequence.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added can be derived from, for

example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold. Buchman and Berg, *Mol. Cell Biol.* 8: 4395-4405 (1988); Callis *et al.*, *Genes Dev.* 1: 1183-1200 (1987). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. See generally, *The Maize Handbook*, Chapter 116, Freeling and Walbot, Eds., Springer, New York (1994).

The vector comprising the sequences from a polynucleotide of the present invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. Usually, the selectable marker gene will encode antibiotic resistance, with suitable genes including genes coding for resistance to the antibiotic spectinomycin (e.g., the *aada* gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance, genes coding for resistance to herbicides which act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides which act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the *bar* gene), or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS gene encodes resistance to the herbicide chlorsulfuron.

Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers *et al.*, *Meth. in Enzymol.*, 153:253-277 (1987). These vectors are plant integrating vectors in that on transformation, the vectors integrate a portion of vector DNA into the genome of the host plant. Exemplary A.

tumefaciens vectors useful herein are plasmids pKYLX6 and pKYLX7 of Schardl *et al.*, *Gene*, 61:1-11 (1987) and Berger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 86:8402-8406 (1989). Another useful vector herein is plasmid pBI101.2 that is available from Clontech Laboratories, Inc. (Palo Alto, CA).

5 A polynucleotide of the present invention can be expressed in either sense or anti-sense orientation as desired. It will be appreciated that control of gene expression in either sense or anti-sense orientation can have a direct impact on the observable plant characteristics. Antisense technology can be conveniently used to gene expression in plants. To accomplish this, a nucleic acid segment from the desired gene is cloned and
10 operably linked to a promoter such that the anti-sense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been shown that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see, e.g., Sheehy *et al.*, *Proc. Nat'l. Acad. Sci. (USA)* 85: 8805-8809 (1988); and Hiatt *et al.*, U.S. Patent No. 4,801,340.
15

~~Another method of suppression is sense suppression. Introduction of nucleic acid~~
configured in the sense orientation has been shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli *et al.*, *The Plant Cell* 2: 279-289
20 (1990) and U.S. Patent No. 5,034,323.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of plant genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is
25 not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. The design and use of target RNA-specific ribozymes is described in Haseloff *et al.*, *Nature* 334: 585-591 (1988).

30 A variety of cross-linking agents, alkylating agents and radical generating species as pendant groups on polynucleotides of the present invention can be used to bind, label, detect, and/or cleave nucleic acids. For example, Vlassov, V. V., *et al.*, *Nucleic Acids Res* (1986) 14:4065-4076, describe covalent bonding of a single-stranded DNA fragment

with alkylating derivatives of nucleotides complementary to target sequences. A report of similar work by the same group is that by Knorre, D. G., *et al.*, *Biochimie* (1985) 67:785-789. Iverson and Dervan also showed sequence-specific cleavage of single-stranded DNA mediated by incorporation of a modified nucleotide which was capable of activating cleavage (*J Am Chem Soc* (1987) 109:1241-1243). Meyer, R. B., *et al.*, *J Am Chem Soc* (1989) 111:8517-8519, effect covalent crosslinking to a target nucleotide using an alkylating agent complementary to the single-stranded target nucleotide sequence. A photoactivated crosslinking to single-stranded oligonucleotides mediated by psoralen was disclosed by Lee, B. L., *et al.*, *Biochemistry* (1988) 27:3197-3203. Use of crosslinking in triple-helix forming probes was also disclosed by Home, *et al.*, *J Am Chem Soc* (1990) 112:2435-2437. Use of N4, N4-ethanocytosine as an alkylating agent to crosslink to single-stranded oligonucleotides has also been described by Webb and Matteucci, *J Am Chem Soc* (1986) 108:2764-2765; *Nucleic Acids Res* (1986) 14:7661-7674; Feteritz *et al.*, *J. Am. Chem. Soc.* 113:4000 (1991). Various compounds to bind, detect, label, and/or cleave nucleic acids are known in the art. See, for example, U.S. Patent Nos. 5,543,507; 5,672,593; 5,484,908; 5,256,648; and, 5,681,941.

Proteins

The isolated proteins of the present invention comprise a polypeptide having at least 10 amino acids encoded by any one of the polynucleotides of the present invention as discussed more fully, *supra*, or polypeptides which are conservatively modified variants thereof. Exemplary polypeptide sequences are provided in SEQ ID NOS: 2, 4, 6, 8, 10, and 12. The proteins of the present invention or variants thereof can comprise any number of contiguous amino acid residues from a polypeptide of the present invention, wherein that number is selected from the group of integers consisting of from 10 to the number of residues in a full-length Rad6 polypeptide. Optionally, this subsequence of contiguous amino acids is at least 15, 20, 25, 30, 35, or 40 amino acids in length, often at least 50, 60, 70, 80, or 90 amino acids in length. Further, the number of such subsequences can be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5.

The present invention further provides a protein comprising a polypeptide having a specified sequence identity with a polypeptide of the present invention. The percentage of sequence identity is an integer selected from the group consisting of from 60 to 99.

Exemplary sequence identity values include 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, and 99%.

As those of skill will appreciate, the present invention includes catalytically active polypeptides of the present invention (i.e., enzymes). Catalytically active polypeptides have a specific activity of at least 20%, 30%, or 40%, and preferably at least 50%, 60%, or 70%, and most preferably at least 80%, 90%, or 95% that of the native (non-synthetic), endogenous polypeptide. Further, the substrate specificity (k_{cat}/K_m) is optionally substantially similar to the native (non-synthetic), endogenous polypeptide.

Typically, the K_m will be at least 30%, 40%, or 50%, that of the native (non-synthetic), endogenous polypeptide; and more preferably at least 60%, 70%, 80%, or 90%. Methods of assaying and quantifying measures of enzymatic activity and substrate specificity (k_{cat}/K_m), are well known to those of skill in the art.

Generally, the proteins of the present invention will, when presented as an immunogen, elicit production of an antibody specifically reactive to a polypeptide of the present invention encoded by a polynucleotide of the present invention as described,

~~*supra*. Exemplary polypeptides include those which are full-length, such as those disclosed in SEQ ID NOS: 2, 4, 6, 8, 10, and 12. Further, the proteins of the present invention will not bind to antisera raised against a polypeptide of the present invention which has been fully immunosorbed with the same polypeptide. Immunoassays for determining binding are well known to those of skill in the art. A preferred immunoassay is a competitive immunoassay as discussed, *infra*. Thus, the proteins of the present invention can be employed as immunogens for constructing antibodies immunoreactive to a protein of the present invention for such exemplary utilities as immunoassays or protein purification techniques.~~

Expression of Proteins in Host Cells

Using the nucleic acids of the present invention, one may express a protein of the present invention in a recombinantly engineered cell such as bacteria, yeast, insect, mammalian, or preferably plant cells. The cells produce the protein in a non-natural condition (e.g., in quantity, composition, location, and/or time), because they have been genetically altered through human intervention to do so.

It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of a nucleic acid encoding a protein of the

present invention. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

In brief summary, the expression of isolated nucleic acids encoding a protein of the present invention will typically be achieved by operably linking, for example, the DNA or cDNA to a promoter (which is either constitutive or inducible), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of the DNA encoding a protein of the present invention. To obtain high level expression of a cloned gene, it is desirable to construct expression vectors which contain, at the minimum, a strong promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. One of skill would recognize that modifications can be made to a protein of the present invention without diminishing its biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of the targeting molecule into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (*e.g.*, poly His) placed on either terminus to create conveniently located restriction sites or termination codons or purification sequences.

A. Expression in Prokaryotes

Prokaryotic cells may be used as hosts for expression. Prokaryotes most frequently are represented by various strains of *E. coli*; however, other microbial strains may also be used. Commonly used prokaryotic control sequences which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta lactamase (penicillinase) and lactose (lac) promoter systems (Chang *et al.*, *Nature* 198:1056 (1977)), the tryptophan (trp) promoter system (Goeddel *et al.*, *Nucleic Acids Res.* 8:4057 (1980)) and the lambda derived P_L promoter and N-gene ribosome binding site (Shimatake *et al.*, *Nature* 292:128 (1981)). The inclusion of selection markers in DNA vectors transfected in *E. coli* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol.

The vector is selected to allow introduction into the appropriate host cell.

Bacterial vectors are typically of plasmid or phage origin. Appropriate bacterial cells are infected with phage vector particles or transfected with naked phage vector DNA. If a plasmid vector is used, the bacterial cells are transfected with the plasmid vector DNA.

5 Expression systems for expressing a protein of the present invention are available using *Bacillus sp.* and *Salmonella* (Palva, *et al.*, *Gene* 22: 229-235 (1983); Mosbach, *et al.*, *Nature* 302: 543-545 (1983)).

B. Expression in Eukaryotes

10 A variety of eukaryotic expression systems such as yeast, insect cell lines, plant and mammalian cells, are known to those of skill in the art. As explained briefly below, a of the present invention can be expressed in these eukaryotic systems. In some embodiments, transformed/transfected plant cells, as discussed *infra*, are employed as expression systems for production of the proteins of the instant invention.

15 Synthesis of heterologous proteins in yeast is well known. Sherman, F., *et al.*, *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory (1982) is a well recognized work describing the various methods available to produce the protein in yeast. Two widely utilized yeast for production of eukaryotic proteins are *Saccharomyces cerevisiae* and *Pichia pastoris*. Vectors, strains, and protocols for expression in *Saccharomyces*
20 and *Pichia* are known in the art and available from commercial suppliers (e.g., Invitrogen). Suitable vectors usually have expression control sequences, such as promoters, including 3-phosphoglycerate kinase or alcohol oxidase, and an origin of replication, termination sequences and the like as desired.

A protein of the present invention, once expressed, can be isolated from yeast by
25 lysing the cells and applying standard protein isolation techniques to the lysates. The monitoring of the purification process can be accomplished by using Western blot techniques or radioimmunoassay of other standard immunoassay techniques.

The sequences encoding proteins of the present invention can also be ligated to various expression vectors for use in transfecting cell cultures of, for instance,
30 mammalian, insect, or plant origin. Illustrative of cell cultures useful for the production of the peptides are mammalian cells. Mammalian cell systems often will be in the form of monolayers of cells although mammalian cell suspensions may also be used. A number of suitable host cell lines capable of expressing intact proteins have been

developed in the art, and include the HEK293, BHK21, and CHO cell lines. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a promoter (*e.g.*, the CMV promoter, a HSV *tk* promoter or *pgk* (phosphoglycerate kinase) promoter), an enhancer (Queen *et al.*, *Immunol. Rev.* 89: 49 (1986)), and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites (*e.g.*, an SV40 large T Ag poly A addition site), and transcriptional terminator sequences. Other animal cells useful for production of proteins of the present invention are available, for instance, from the American Type Culture Collection Catalogue of Cell Lines and Hybridomas (7th edition, 1992).

Appropriate vectors for expressing proteins of the present invention in insect cells are usually derived from the SF9 baculovirus. Suitable insect cell lines include mosquito larvae, silkworm, armyworm, moth and *Drosophila* cell lines such as a Schneider cell line (See Schneider, *J. Embryol. Exp. Morphol.* 27: 353-365 (1987)).

As with yeast, when higher animal or plant host cells are employed, polyadenylation or transcription terminator sequences are typically incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VP1 intron from SV40 (Sprague, *et al.*, *J. Virol.* 45: 773-781 (1983)). Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors. Saveria-Campo, M., Bovine Papilloma Virus DNA a Eukaryotic Cloning Vector in *DNA Cloning Vol. II a Practical Approach*, D.M. Glover, Ed., IRL Press, Arlington, Virginia pp. 213-238 (1985).

Transfection/Transformation of Cells

The method of transformation/transfection is not critical to the instant invention; various methods of transformation or transfection are currently available. As newer methods are available to transform crops or other host cells they may be directly applied. Accordingly, a wide variety of methods have been developed to insert a DNA sequence into the genome of a host cell to obtain the transcription and/or translation of the sequence to effect phenotypic changes in the organism. Thus, any method which provides for efficient transformation/transfection may be employed.

A. Plant Transformation

A DNA sequence coding for the desired polynucleotide of the present invention, for example a cDNA or a genomic sequence encoding a full length protein, will be used to construct a recombinant expression cassette which can be introduced into the desired plant.

Isolated nucleic acid acids of the present invention can be introduced into plants according techniques known in the art. Generally, recombinant expression cassettes as described above and suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical, scientific, and patent literature. See, for example, Weising *et al.*, *Ann. Rev. Genet.* 22: 421-477 (1988). For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation, PEG poration, particle bombardment, silicon fiber delivery, or microinjection of plant cell protoplasts or embryogenic callus. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria. See also, Hiei, *et al.*, U. S. patent no. 5,591,616.

The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al.*, *Embo J.* 3: 2717-2722 (1984). Electroporation techniques are described in Fromm *et al.*, *Proc. Natl. Acad. Sci.* 82: 5824 (1985). Ballistic transformation techniques are described in Klein *et al.*, *Nature* 327: 70-73 (1987).

Agrobacterium tumefaciens-mediated transformation techniques are well described in the scientific literature. See, for example Horsch *et al.*, *Science* 233: 496-498 (1984), and Fraley *et al.*, *Proc. Natl. Acad. Sci.* 80: 4803 (1983). Although *Agrobacterium* is useful primarily in dicots, certain monocots can be transformed by *Agrobacterium*. For instance, *Agrobacterium* transformation of maize is described in U.S. Patent No. 5,550,318.

Other methods of transfection or transformation include (1) *Agrobacterium rhizogenes*-mediated transformation (see, e.g., Lichtenstein and Fuller In: Genetic Engineering, vol. 6, PWJ Rigby, Ed., London, Academic Press, 1987; and Lichtenstein,

C. P., and Draper, J., In: DNA Cloning, Vol. II, D. M. Glover, Ed., Oxford, IRI Press, 1985), Application PCT/US87/02512 (WO 88/02405 published Apr. 7, 1988) describes the use of *A. rhizogenes* strain A4 and its Ri plasmid along with *A. tumefaciens* vectors pARC8 or pARC16 (2) liposome-mediated DNA uptake (see, e.g., Freeman *et al.*, Plant Cell Physiol. 25: 1353, 1984), (3) the vortexing method (see, e.g., Kindle, Proc. Natl. Acad. Sci., USA 87: 1228, (1990).

DNA can also be introduced into plants by direct DNA transfer into pollen as described by Zhou *et al.*, Methods in Enzymology, 101:433 (1983); D. Hess, Intern Rev. Cytol., 107:367 (1987); Luo *et al.*, Plant Mol. Biol. Reporter, 6:165 (1988).

- 10 Expression of polypeptide coding genes can be obtained by injection of the DNA into reproductive organs of a plant as described by Pena *et al.*, Nature, 325:274 (1987). DNA can also be injected directly into the cells of immature embryos and the rehydration of desiccated embryos as described by Neuhaus *et al.*, Theor. Appl. Genet., 75:30 (1987); and Benbrook *et al.*, in Proceedings Bio Expo 1986, Butterworth, 15 Stoneham, Mass., pp. 27-54 (1986). A variety of plant viruses that can be employed as vectors are known in the art and include cauliflower mosaic virus (CaMV), geminivirus, brome mosaic virus, and tobacco mosaic virus.

B. Transfection of Prokaryotes, Lower Eukaryotes, and Animal Cells

- 20 Animal and lower eukaryotic (e.g., yeast) host cells are competent or rendered competent for transfection by various means. There are several well-known methods of introducing DNA into animal cells. These include: calcium phosphate precipitation, fusion of the recipient cells with bacterial protoplasts containing the DNA, treatment of the recipient cells with liposomes containing the DNA, DEAE dextran, electroporation, 25 biolistics, and micro-injection of the DNA directly into the cells. The transfected cells are cultured by means well known in the art. Kuchler, R.J., *Biochemical Methods in Cell Culture and Virology*, Dowden, Hutchinson and Ross, Inc. (1977).

Synthesis of Proteins

- 30 The proteins of the present invention can be constructed using non-cellular synthetic methods. Solid phase synthesis of proteins of less than about 50 amino acids in length may be accomplished by attaching the C-terminal amino acid of the sequence to an insoluble support followed by sequential addition of the remaining amino acids in the

sequence. Techniques for solid phase synthesis are described by Barany and Merrifield, Solid-Phase Peptide Synthesis, pp. 3-284 in *The Peptides: Analysis, Synthesis, Biology. Vol. 2: Special Methods in Peptide Synthesis, Part A.*; Merrifield, *et al.*, *J. Am. Chem. Soc.* 85: 2149-2156 (1963), and Stewart *et al.*, *Solid Phase Peptide Synthesis, 2nd ed.*, Pierce Chem. Co., Rockford, Ill. (1984). Proteins of greater length may be synthesized by condensation of the amino and carboxy termini of shorter fragments. Methods of forming peptide bonds by activation of a carboxy terminal end (e.g., by the use of the coupling reagent N,N'-dicyclohexylcarbodiimide)) is known to those of skill.

10 Purification of Proteins

The proteins of the present invention may be purified by standard techniques well known to those of skill in the art. Recombinantly produced proteins of the present invention can be directly expressed or expressed as a fusion protein. The recombinant protein is purified by a combination of cell lysis (e.g., sonication, French press) and affinity chromatography. For fusion products, subsequent digestion of the fusion protein with an appropriate proteolytic enzyme releases the desired recombinant protein.

The proteins of this invention, recombinant or synthetic, may be purified to substantial purity by standard techniques well known in the art, including detergent solubilization, selective precipitation with such substances as ammonium sulfate, column chromatography, immunopurification methods, and others. See, for instance, R. Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag: New York (1982); Deutscher, *Guide to Protein Purification*, Academic Press (1990). For example, antibodies may be raised to the proteins as described herein. Purification from *E. coli* can be achieved following procedures described in U.S. Patent No. 4,511,503. The protein may then be isolated from cells expressing the protein and further purified by standard protein chemistry techniques as described herein. Detection of the expressed protein is achieved by methods known in the art and include, for example, radioimmunoassays, Western blotting techniques or immunoprecipitation.

30 Transgenic Plant Regeneration

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype. Such regeneration techniques often rely on manipulation of certain

phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with a polynucleotide of the present invention. For transformation and regeneration of maize see, Gordon-Kamm *et al.*, *The Plant Cell*, 2:603-618 (1990).

- 5 Plants cells transformed with a plant expression vector can be regenerated, e.g., from single cells, callus tissue or leaf discs according to standard plant tissue culture techniques. It is well known in the art that various cells, tissues, and organs from almost any plant can be successfully cultured to regenerate an entire plant. Plant regeneration from cultured protoplasts is described in Evans *et al.*, *Protoplasts Isolation and Culture*,
10 *Handbook of Plant Cell Culture*, Macmillan Publishing Company, New York, pp. 124-176 (1983); and Binding, *Regeneration of Plants, Plant Protoplasts*, CRC Press, Boca Raton, pp. 21-73 (1985).

- The regeneration of plants containing the foreign gene introduced by *Agrobacterium* from leaf explants can be achieved as described by Horsch *et al.*,
15 *Science*, 227:1229-1231 (1985). In this procedure, transformants are grown in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant species being transformed as described by Fraley *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 80:4803 (1983). This procedure typically produces shoots within two to four weeks and these transformant shoots are then transferred to an appropriate root-inducing
20 medium containing the selective agent and an antibiotic to prevent bacterial growth. Transgenic plants of the present invention may be fertile or sterile.

- Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al.*, *Ann. Rev. of Plant Phys.* 38: 467-486 (1987). The regeneration of plants from either single plant
25 protoplasts or various explants is well known in the art. See, for example, *Methods for Plant Molecular Biology*, A. Weissbach and H. Weissbach, eds., Academic Press, Inc., San Diego, Calif. (1988). This regeneration and growth process includes the steps of selection of transformant cells and shoots, rooting the transformant shoots and growth of the plantlets in soil. For maize cell culture and regeneration see generally, *The Maize*
30 *Handbook*, Freeling and Walbot, Eds., Springer, New York (1994); *Corn and Corn Improvement*, 3rd edition, Sprague and Dudley Eds., American Society of Agronomy, Madison, Wisconsin (1988).

One of skill will recognize that after the recombinant expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

5 In vegetatively propagated crops, mature transgenic plants can be propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. Selection of desirable transgenics is made and new varieties are obtained and propagated vegetatively for commercial use. In seed propagated crops, mature transgenic plants can be self crossed to produce a homozygous inbred plant. The inbred plant produces
10 seed containing the newly introduced heterologous nucleic acid. These seeds can be grown to produce plants that would produce the selected phenotype, (e.g., altered Rad6 content or composition).

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are included in the invention, provided that these parts
15 comprise cells comprising the isolated nucleic acid of the present invention. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced nucleic acid sequences.

Transgenic plants expressing the selectable marker can be screened for transmission of the nucleic acid of the present invention by, for example, standard
20 immunoblot and DNA detection techniques. Transgenic lines are also typically evaluated on levels of expression of the heterologous nucleic acid. Expression at the RNA level can be determined initially to identify and quantitate expression-positive plants. Standard techniques for RNA analysis can be employed and include PCR amplification assays using oligonucleotide primers designed to amplify only the heterologous RNA templates
25 and solution hybridization assays using heterologous nucleic acid-specific probes. The RNA-positive plants can then analyzed for protein expression by Western immunoblot analysis using the specifically reactive antibodies of the present invention. In addition, *in situ* hybridization and immunocytochemistry according to standard protocols can be done using heterologous nucleic acid specific polynucleotide probes and antibodies,
30 respectively, to localize sites of expression within transgenic tissue. Generally, a number of transgenic lines are usually screened for the incorporated nucleic acid to identify and select plants with the most appropriate expression profiles.

A preferred embodiment is a transgenic plant that is homozygous for the added heterologous nucleic acid; i.e., a transgenic plant that contains two added nucleic acid sequences, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) a heterozygous transgenic plant that contains a single added heterologous nucleic acid, germinating some of the seed produced and analyzing the resulting plants produced for altered Rad6 expression relative to a control plant (i.e., native, non-transgenic). Back-crossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated.

10 Modulating Rad6 Levels

The present invention further provides a method for modulating (i.e., increasing or decreasing) Rad6 levels in a plant or part thereof. Modulation can be effected by increasing or decreasing the total amount of Rad6 (i.e., its content) and/or the ratio of various Rad6 proteins (i.e., its composition) in the plant. The method comprises transforming a plant cell with a recombinant expression cassette comprising a polynucleotide of the present invention as described above to obtain a transformed plant cell, growing the transformed plant cell under plant forming conditions, and inducing expression of a polynucleotide of the present invention in the plant for a time sufficient to modulate Rad6 content and/or composition in the plant or plant part.

20 In some embodiments, Rad6 in a plant may be modulated by altering, *in vivo* or *in vitro*, the promoter of a non-isolated Rad6 gene to up- or down-regulate gene expression. In some embodiments, the coding regions of native Rad6 genes can be altered via substitution, addition, insertion, or deletion to decrease activity of the encoded enzyme. See, e.g., Kmiec, U.S. Patent 5,565,350; Zarling *et al.*,
25 PCT/US93/03868. And in some embodiments, an isolated nucleic acid (e.g., a vector) comprising a promoter sequence is transfected into a plant cell. Subsequently, a plant cell comprising the promoter operably linked to a polynucleotide of the present invention is selected for by means known to those of skill in the art such as, but not limited to, Southern blot, DNA sequencing, or PCR analysis using primers specific to the promoter
30 and to the gene and detecting amplicons produced therefrom. A plant or plant part altered or modified by the foregoing embodiments is grown under plant forming conditions for a time sufficient to modulate Rad6 content and/or composition in the plant. Plant forming conditions are well known in the art and discussed briefly, *supra*.

In general, content or composition is increased or decreased by at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% relative to a native control plant, plant part, or cell lacking the aforementioned recombinant expression cassette. Modulation in the present invention may occur during and/or subsequent to growth of the plant to the desired stage of development. Modulating nucleic acid expression temporally and/or in particular tissues can be controlled by employing the appropriate promoter operably linked to a polynucleotide of the present invention in, for example, sense or antisense orientation as discussed in greater detail, *supra*. Induction of expression of a polynucleotide of the present invention can also be controlled by exogenous administration of an effective amount of inducing compound. Inducible promoters and inducing compounds which activate expression from these promoters are well known in the art. In preferred embodiments, Rad6 is modulated in monocots, particularly maize.

Molecular Markers

The present invention provides a method of genotyping a plant comprising a polynucleotide of the present invention. Preferably, the plant is a monocot, such as maize or sorghum. Genotyping provides a means of distinguishing homologs of a chromosome pair and can be used to differentiate segregants in a plant population. Molecular marker methods can be used for phylogenetic studies, characterizing genetic relationships among crop varieties, identifying crosses or somatic hybrids, localizing chromosomal segments affecting monogenic traits, map based cloning, and the study of quantitative inheritance. See, e.g., *Plant Molecular Biology: A Laboratory Manual*, Chapter 7, Clark, Ed., Springer-Verlag, Berlin (1997). For molecular marker methods, see generally, *The DNA Revolution* by Andrew H. Paterson 1996 (Chapter 2) in: *Genome Mapping in Plants* (ed. Andrew H. Paterson) by Academic Press/R. G. Landis Company, Austin, Texas, pp.7-21.

The particular method of genotyping in the present invention may employ any number of molecular marker analytic techniques such as, but not limited to, restriction fragment length polymorphisms (RFLPs). RFLPs are the product of allelic differences between DNA restriction fragments caused by nucleotide sequence variability. As is well known to those of skill in the art, RFLPs are typically detected by extraction of genomic DNA and digestion with a restriction enzyme. Generally, the resulting fragments are separated according to size and hybridized with a probe; single copy

probes are preferred. Restriction fragments from homologous chromosomes are revealed. Differences in fragment size among alleles represent an RFLP. Thus, the present invention further provides a means to follow segregation of a Rad6 gene or nucleic acid of the present invention as well as chromosomal sequences genetically linked to these genes or nucleic acids using such techniques as RFLP analysis. Linked chromosomal sequences are within 50 centiMorgans (cM), often within 40 or 30 cM, preferably within 20 or 10 cM, more preferably within 5, 3, 2, or 1 cM of a Rad6 gene.

In the present invention, the nucleic acid probes employed for molecular marker mapping of plant nuclear genomes selectively hybridize, under selective hybridization conditions, to a gene encoding a polynucleotide of the present invention. In preferred embodiments, the probes are selected from polynucleotides of the present invention. Typically, these probes are cDNA probes or *Pst I* genomic clones. The length of the probes is discussed in greater detail, *supra*, but are typically at least 15 bases in length, more preferably at least 20, 25, 30, 35, 40, or 50 bases in length. Generally, however, the probes are less than about 1 kilobase in length. Preferably, the probes are single copy probes that hybridize to a unique locus in a haploid chromosome complement. Some exemplary restriction enzymes employed in RFLP mapping are EcoRI, EcoRV, and SstI. As used herein the term "restriction enzyme" includes reference to a composition that recognizes and, alone or in conjunction with another composition, cleaves at a specific nucleotide sequence.

The method of detecting an RFLP comprises the steps of (a) digesting genomic DNA of a plant with a restriction enzyme; (b) hybridizing a nucleic acid probe, under selective hybridization conditions, to a sequence of a polynucleotide of the present of said genomic DNA; (c) detecting therefrom a RFLP. Other methods of differentiating polymorphic (allelic) variants of polynucleotides of the present invention can be had by utilizing molecular marker techniques well known to those of skill in the art including such techniques as: 1) single stranded conformation analysis (SSCP); 2) denaturing gradient gel electrophoresis (DGGE); 3) RNase protection assays; 4) allele-specific oligonucleotides (ASOs); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein; and 6) allele-specific PCR. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE); heteroduplex analysis (HA); and chemical mismatch cleavage (CMC). Exemplary polymorphic variants are provided in

Table I, *supra*. Thus, the present invention further provides a method of genotyping comprising the steps of contacting, under stringent hybridization conditions, a sample suspected of comprising a polynucleotide of the present invention with a nucleic acid probe. Generally, the sample is a plant sample; preferably, a sample suspected of comprising a maize polynucleotide of the present invention (e.g., gene, mRNA). The nucleic acid probe selectively hybridizes, under stringent conditions, to a subsequence of a polynucleotide of the present invention comprising a polymorphic marker. Selective hybridization of the nucleic acid probe to the polymorphic marker nucleic acid sequence yields a hybridization complex. Detection of the hybridization complex indicates the presence of that polymorphic marker in the sample. In preferred embodiments, the nucleic acid probe comprises a polynucleotide of the present invention.

UTR's and Codon Preference

In general, translational efficiency has been found to be regulated by specific sequence elements in the 5' non-coding or untranslated region (5' UTR) of the RNA. Positive sequence motifs include translational initiation consensus sequences (Kozak, *Nucleic Acids Res.* 15:8125 (1987)) and the 7-methylguanosine cap structure (Drummond *et al.*, *Nucleic Acids Res.* 13:7375 (1985)). Negative elements include stable intramolecular 5' UTR stem-loop structures (Muesing *et al.*, *Cell* 48:691 (1987)) and AUG sequences or short open reading frames preceded by an appropriate AUG in the 5' UTR (Kozak, *supra*, Rao *et al.*, *Mol. and Cell. Biol.* 8:284 (1988)). Accordingly, the present invention provides 5' and/or 3' UTR regions for modulation of translation of heterologous coding sequences.

Further, the polypeptide-encoding segments of the polynucleotides of the present invention can be modified to alter codon usage. Altered codon usage can be employed to alter translational efficiency and/or to optimize the coding sequence for expression in a desired host or to optimize the codon usage in a heterologous sequence for expression in maize. Codon usage in the coding regions of the polynucleotides of the present invention can be analyzed statistically using commercially available software packages such as "Codon Preference" available from the University of Wisconsin Genetics Computer Group (see Devereaux *et al.*, *Nucleic Acids Res.* 12: 387-395 (1984)) or MacVector 4.1 (Eastman Kodak Co., New Haven, Conn.). Thus, the present invention provides a codon usage frequency characteristic of the coding region of at least one of

the polynucleotides of the present invention. The number of polynucleotides that can be used to determine a codon usage frequency can be any integer from 1 to the number of polynucleotides of the present invention as provided herein. Optionally, the polynucleotides will be full-length sequences. An exemplary number of sequences for statistical analysis can be at least 1, 5, 10, 20, 50, or 100.

Sequence Shuffling

The present invention provides methods for sequence shuffling using polynucleotides of the present invention, and compositions resulting therefrom.

Sequence shuffling is described in PCT publication No. WO 97/20078. See also, Zhang, J.-H., *et al. Proc. Natl. Acad. Sci. USA* 94:4504-4509 (1997). Generally, sequence shuffling provides a means for generating libraries of polynucleotides having a desired characteristic which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides that comprise sequence regions which have substantial sequence identity and can be homologously recombined in vitro or in vivo. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property. In some embodiments, the selected characteristic will be a decreased K_m and/or increased K_{cat} over the wild-type protein as provided herein. In other embodiments, a protein or polynucleotide generated from sequence shuffling will have a ligand binding affinity greater than the non-shuffled wild-type polynucleotide. The increase in such properties can be at least 110%, 120%, 130%, 140% or at least 150% of the wild-type value.

Assays for Compounds that Modulate Enzymatic Activity or Expression

The present invention also provides means for identifying compounds that bind to (e.g., substrates), and/or increase or decrease (i.e., modulate) the enzymatic activity of,

catalytically active polypeptides of the present invention. The method comprises contacting a polypeptide of the present invention with a compound whose ability to bind to or modulate enzyme activity is to be determined. The polypeptide employed will have at least 20%, preferably at least 30% or 40%, more preferably at least 50% or 60%, and most preferably at least 70% or 80% of the specific activity of the native, full-length Rad6 polypeptide (e.g., enzyme). Generally, the polypeptide will be present in a range sufficient to determine the effect of the compound, typically about 1 nM to 10 μ M. Likewise, the compound will be present in a concentration of from about 1 nM to 10 μ M. Those of skill will understand that such factors as enzyme concentration, ligand concentrations (i.e., substrates, products, inhibitors, activators), pH, ionic strength, and temperature will be controlled so as to obtain useful kinetic data and determine the presence or absence of a compound that binds or modulates polypeptide activity. Methods of measuring enzyme kinetics is well known in the art. See, e.g., Segel, *Biochemical Calculations*, 2nd ed., John Wiley and Sons, New York (1976).

Although the present invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims:

Example 1

This example describes the construction cDNA libraries.

Libraries

Table 1 provides a summary of the Rad6 clones isolated, and the tissue type, developmental stage, and tissue treatment used to construct the libraries from which the clones were isolated.

TABLE 1

<u>Clone ID</u>	<u>Clone Name</u>	<u>SEQ ID NO:</u>	<u>Library No., Tissue, Developmental Stage and Treatment</u>
CFTAA48	zmUBC-1	n/a	P0040 Tassel morning before floral transition
CIABG54	zmUBC-2	1	P0103 Tassel shoots, screened
CMTAA60	zmRAD6-1	11	P0036 Early meiosis tassel

	CRWBG76	zmRAD6-2	9	P0037 Root worm infested root
	CTBAF50	zmRAD6-3	7	P0099 Tassel Apical Meristem
	CTSBC96	zmUBC-3	n/a	P0016 Tassel shoots
	CTSN61	zmUBC-4	n/a	P0016 Tassel shoots
5	CTSCK60	zmUBC-5	5	P0016 Tassel shoots
	CTSC071	zmUBC-6	n/a	P0016 Tassel shoots
	CXMAG13	zmUBC-7	3	P0061 Anther prophase

Total RNA Isolation

10 Total RNA was isolated from corn tissues with TRIzol Reagent (Life Technology Inc. Gaithersburg, MD) using a modification of the guanidine isothiocyanate/acid-phenol procedure described by Chomczynski and Sacchi (Chomczynski, P., and Sacchi, N. *Anal. Biochem.* 162, 156 (1987)). In brief, plant tissue samples were pulverized in liquid nitrogen before the addition of the TRIzol Reagent, and then were further

15 homogenized with a mortar and pestle. Addition of chloroform followed by centrifugation was conducted for separation of an aqueous phase and an organic phase. The total RNA was recovered by precipitation with isopropyl alcohol from the aqueous phase.

20 Poly(A)+ RNA Isolation

The selection of poly(A)+ RNA from total RNA was performed using PolyATtract system (Promega Corporation, Madison, WI). In brief, biotinylated oligo(dT) primers were used to hybridize to the 3' poly(A) tails on mRNA. The hybrids were captured using streptavidin coupled to paramagnetic particles and a magnetic

25 separation stand. The mRNA was washed at high stringent condition and eluted by RNase-free deionized water.

cDNA Library Construction

cDNA synthesis was performed and unidirectional cDNA libraries were

30 constructed using the SuperScript Plasmid System (Life Technology Inc. Gaithersburg, MD). The first strand of cDNA was synthesized by priming an oligo(dT) primer containing a Not I site. The reaction was catalyzed by SuperScript Reverse Transcriptase II at 45°C. The second strand of cDNA was labeled with alpha-³²P-dCTP and a portion of the reaction was analyzed by agarose gel electrophoresis to determine

cDNA sizes. cDNA molecules smaller than 500 base pairs and unligated adapters were removed by Sephacryl-S400 chromatography. The selected cDNA molecules were ligated into pSPORT1 vector in between of Not I and Sal I sites.

5 Example 2

This example describes cDNA sequencing and library subtraction.

Sequencing Template Preparation

Individual colonies were picked and DNA was prepared either by PCR with M13 forward primers and M13 reverse primers, or by plasmid isolation. All the cDNA
10 clones were sequenced using M13 reverse primers.

Q-bot Subtraction Procedure

cDNA libraries subjected to the subtraction procedure were plated out on 22 x 22 cm² agar plate at density of about 3,000 colonies per plate. The plates were incubated in a 37°C incubator for 12-24 hours. Colonies were picked into 384-well plates by a robot
15 colony picker, Q-bot (GENETIX Limited). These plates were incubated overnight at 37°C.

Once sufficient colonies were picked, they were pinned onto 22 x 22 cm² nylon membranes using Q-bot. Each membrane contained 9,216 colonies or 36,864 colonies. These membranes were placed onto agar plate with appropriate antibiotic. The plates
20 were incubated at 37°C for overnight.

After colonies were recovered on the second day, these filters were placed on filter paper prewetted with denaturing solution for four minutes, then were incubated on top of a boiling water bath for additional four minutes. The filters were then placed on filter paper prewetted with neutralizing solution for four minutes. After excess solution
25 was removed by placing the filters on dry filter papers for one minute, the colony side of the filters were place into Proteinase K solution, incubated at 37°C for 40-50 minutes. The filters were placed on dry filter papers to dry overnight. DNA was then cross-linked to nylon membrane by UV light treatment.

Colony hybridization was conducted as described by Sambrook, J., Fritsch, E.F. and
30 Maniatis, T., (in Molecular Cloning: A laboratory Manual, 2nd Edition). The following probes were used in colony hybridization:

1. First strand cDNA from the same tissue as the library was made from to remove the most redundant clones.

2. 48-192 most redundant cDNA clones from the same library based on previous sequencing data.
3. 192 most redundant cDNA clones in the entire corn sequence database.
4. A Sal-A20 oligo nucleotide: TCG ACC CAC GCG TCC GAA AAA AAA AAA
- 5 AAA AAA AAA, listed in SEQ ID NO: 13, removes clones containing a poly A tail but no cDNA.
5. cDNA clones derived from rRNA.

The image of the autoradiography was scanned into computer and the signal intensity and cold colony addresses of each colony was analyzed. Re-arraying of cold-colonies from 384 well plates to 96 well plates was conducted using Q-bot.

Table 2 is a pileup of the proteins of the present invention including a yeast Rad6 sequence (ScRad6, GenBank accession number A21906).

TABLE 2

Amino acid sequence comparison of maize and yeast RAD6

		1		50
	zmUBC2 (SEQ ID 2)	---MASKRIL	KELKDLQKDP PTSCSAGPAG EDMFHWQATI MGPPDSPYAG	
	zmUBC7 (SEQ ID 4)	---MASKRIL	KELKDLQKDP PTSCSAGPAG EDMFHWQATI MGPPDSPYAG	
20	zmUBC5 (SEQ ID 6)	---MASKRIL	KELKDLQKDP PTSCSAGPVA EDMFYWQATI MGPPDSPYAG	
	zmRAD6-3 (SEQ ID 8)	MSTPARKRLM	RDFKRLQQDP PAGISGAPYD NNIMLWNAVI FGPDDTPWDG	
	zmRAD6-2 (SEQ ID 10)	MSTPARKRLM	RDFKRLQQDP PAGISGAPYD NNIMLWNAVI FGPDDTPWDG	
	zmRAD6-1 (SEQ ID 12)	MSTPARKRLM	RDFKRLMQDP PAGISGAPQD NNIMLWNAVI FGPDDTPWDG	
	ScRAD6 (SEQ ID 14)	MSTPARRRLM	RDFKRMKEDA PPGVSASPLP DNVMVWNAMI IGPADTPYED	
25		51		100
	zmUBC2 (SEQ ID 2)	GVFLVNIHFP	PDYPFKPPKV SFKTKVFHPN INSNGSICLD ILKEQWSPAL	
	zmUBC7 (SEQ ID 4)	GVFLVNIHFP	PDYPFKPPKV SFKTKVFHPN INSNGSICLD ILKEQWSPAL	
	zmUBC5 (SEQ ID 6)	GVFLVTIHFP	PDYPFKPPKV AFKTKVYHPN INSNGSICLD ILKEQWSPAL	
30	zmRAD6-3 (SEQ ID 8)	GTFKLTLOCT	EDYPNKPPTV RFVSRMFHPN IYADGSICLD ILQNQWSPYI	
	zmRAD6-2 (SEQ ID 10)	GTFKLTLOCT	EDYPNKPPTV RFVSRMFHPN IYADGSICLD ILQNKWSPYI	
	zmRAD6-1 (SEQ ID 12)	GTFKLTLOFN	EEYPNKPPTV RFVSRMFHPN IYADGSICLD ILQDQWSPYI	
	ScRAD6 (SEQ ID 14)	GTFRLLEFD	EEYPNKPPTHV KFLSEMFHPN VYANGEICLD ILQNRWTPTY	
35		101		150
	zmUBC2 (SEQ ID 2)	TISKVLLSIC	SLLTDPNPDD PLVPEIAHMY KTRPKYEST ARSWTQKYAM	
	zmUBC7 (SEQ ID 4)	TISKVLLSIC	SLLTDPNPDD PLVPEIAHMY KTRPKYEAT ARSWTQKYAM	
	zmUBC5 (SEQ ID 6)	TVSKVLLSIC	SLLTDPNPDD PLVPEIAHLY KTRVKYEST ARSWTQKYAM	

zmRAD6-3 (SEQ ID 8) DVAAILTSIQ SLLCDPNPNS PANSEAARMF SENKREYNRK VREVVEQSWT
 zmRAD6-2 (SEQ ID 10) DVAAILTSIQ SLLCDPNPNS PANSEAGRMF SENKREYNRK VREVVEQSWT
 zmRAD6-1 (SEQ ID 12) DVDAILTSIQ SLLCDPNPNS PANSEAARMF SENKREYNRK VREVVEQSWT
 ScRAD6 (SEQ ID 14) DVASILTSIQ SLFNDPNPAS PANVEAATLF KDHKSQYVKR VKETVEKSWE

5
 151 172
 zmUBC2 (SEQ ID 2) G*-----
 zmUBC7 (SEQ ID 4) G*-----
 zmUBC5 (SEQ ID 6) G*-----
 10 zmRAD6-3 (SEQ ID 8) AD*-----
 zmRAD6-2 (SEQ ID 10) AD*-----
 zmRAD6-1 (SEQ ID 12) AD*-----
 ScRAD6 (SEQ ID 14) DDMDDMDDDD DDDDDDDDE AD

Example 3

This example describes identification of the gene from a computer homology search.

Gene identities were determined by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1990) J. Mol. Biol. 215:403-410; see also

20 www.ncbi.nlm.nih.gov/BLAST/) searches under default parameters for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences were analyzed
 25 for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm. The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics* 3:266-272) provided by the NCBI. In some cases, the sequencing data
 30 from two or more clones containing overlapping segments of DNA were used to construct contiguous DNA sequences.

A BLAST search of the GenBank (release 110) and EMBL (release 56) databases with the polynucleotide sequences of the present invention revealed the following best hits: for SEQ ID NO. 3, the best hit is accession no. AF034946; for SEQ ID NO. 5, the
 35 best hit is accession no. L29077; for SEQ ID NOS. 7, 9, and 11, the best hit is accession no. M62720.

Table 3 represents comparison of each sequence with itself as well as every other sequence. In the former case, the top number represents total number of amino acids. Each number is expressed as a count and as percentage on different sides of the matrix diagonal. The diagonal shows how many locations have at least one residue for the single sequence, i.e., the sequence length. The top number in each column represents identity and lower number represents identity + conservative substitutions. Thus, when the numbers are expressed as percentages, the top number in each column represents % identity and lower number represents % identity + conservative substitutions. As an example, CIABG51 is 99% identical to CXMAG11 and 100% similar if one considers conservative substitutions.

TABLE 3**Comparative Analysis of Maize RAD6/UBC Sequences and Yeast RAD6**

	ZmUBC2 SEQ ID 2	ZmUBC7 SEQ ID 4	ZmUBC5 SEQ ID 6	ZmRAD6-3 SEQ ID 8	ZmRAD6-2 SEQ ID 10	ZmRAD6-1 SEQ ID 12	ScRAD6 SEQ ID 14
ZmUBC2 SEQ ID 2	148 0	99% 100%	93% 96%	40% 59%	39% 59%	40% 59%	31% 52%
ZmUBC7 SEQ ID 4	147 148	148 0	92% 96%	40% 59%	39% 59%	40% 59%	31% 52%
ZmUBC5 SEQ ID 6	138 143	137 143	148 0	40% 60%	39% 59%	40% 60%	32% 53%
ZmRAD6-3 SEQ ID 8	62 91	62 91	62 91	152 0	98% 99%	95% 96%	55% 73%
ZmRAD6-2 SEQ ID 10	60 90	60 90	60 91	150 151	152 0	94% 96%	55% 73%
ZmRAD6-1 SEQ ID 12	61 91	61 91	61 92	145 147	143 146	152 0	55 73
ScRAD6 SEQ ID 14	55 90	55 90	56 92	96 127	95 126	96 127	172 0

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid comprising a member selected from the group consisting of:
 - a) a polynucleotide encoding a polypeptide selected from the group consisting of
SEQ ID NOS: 4, 6, 8, 10, or 12;
 - 5 b) a polynucleotide having at least 88% sequence identity to SEQ ID NO: 3, at least
81% identity to SEQ ID NO: 5, at least 93% identity to SEQ ID NO: 7, at least
92% identity to SEQ ID NO: 9, or at least 83% identity to SEQ ID NO: 11,
wherein the % sequence identity is based on the complete sequence and is
determined by the GAP program using default parameters;
 - 10 c) a polynucleotide selected from the group consisting of SEQ ID NOS: 3, 5, 7, 9,
and 11;
 - d) a polynucleotide which is complementary to a polynucleotide of (a), (b), or (c);
 - e) a polynucleotide comprising at least 33 contiguous nucleotides from a
polynucleotide of (a), (b), (c), or (d).
- 15 2) A recombinant expression cassette, comprising a member of claim 1 operably linked,
in sense or anti-sense orientation, to a promoter.
- 3) A host cell comprising the recombinant expression cassette of claim 2.
- 20 4) A transgenic plant comprising a recombinant expression cassette of claim 2.
- 5) The transgenic plant of claim 4, wherein said plant is a monocot.
- 25 6) The transgenic plant of claim 4, wherein said plant is a dicot.
- 7) The transgenic plant of claim 4, wherein said plant is selected from the group
consisting of: maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton,
rice, barley, and millet.
- 30 8) A transgenic seed from the transgenic plant of claim 4.
- 9) A method of modulating the level of Rad6 in a plant, comprising:

- (a) introducing into a plant cell a recombinant expression cassette comprising a Rad6 polynucleotide of claim 1 operably linked to a promoter;
- (b) culturing the plant cell under plant cell growing conditions; and
- (c) inducing expression of said polynucleotide for a time sufficient to modulate the level of Rad6 in said plant.

10) The method of claim 9, wherein the plant is maize.

11) An isolated protein comprising a member selected from the group consisting of:

(a) a polypeptide selected from the group consisting of SEQ ID NOS: 4, 6, 8, 10, or 12;

(b) a polypeptide comprising an amino acid sequence which has at least 94% sequence identity to the polypeptide of SEQ ID NO: 6, at least 98% sequence identity to SEQ ID NO: 8, at least 97% identity to SEQ ID NO: 10, or at least 95% identity to SEQ ID NO: 12, wherein the % sequence identity is based on the entire sequence and is determined by the GAP program using default parameters;

(c) and at least one polypeptide encoded by a member of claim 1.

SEQUENCE LISTING

<110> Pioneer Hi-Bred International, Inc.

<120> Maize Rad6 Genes and Uses Thereof

<130> 0883-PCT

<150> 60/096,546

<151> 1998-08-14

<160> 14

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 447

<212> DNA

<213> Zea mays

<220>

<221> CDS

<222> (1) ... (442)

<400> 1

atg gcg tct aag cgc atc ctc aag gag ctg aag gac ctg cag aaa gac	48
Met Ala Ser Lys Arg Ile Leu Lys Glu Leu Lys Asp Leu Gln Lys Asp	
1 5 10 15	
ccc ccc aca tca tgc agt gca ggt cct gct ggt gag gac atg ttt cat	96
Pro Pro Thr Ser Cys Ser Ala Gly Pro Ala Gly Glu Asp Met Phe His	
20 25 30	
tgg caa gca aca att atg gga cca cct gac agt ccc tat gct ggc ggt	144
Trp Gln Ala Thr Ile Met Gly Pro Pro Asp Ser Pro Tyr Ala Gly Gly	
35 40 45	
gtt ttc tta gtg aac att cat ttc ccg cca gat tac ccc ttc aaa cct	192
Val Phe Leu Val Asn Ile His Phe Pro Pro Asp Tyr Pro Phe Lys Pro	
50 55 60	
cca aag gtt tct ttc aag aca aag gtc ttc cac cct aat atc aac agc	240
Pro Lys Val Ser Phe Lys Thr Lys Val Phe His Pro Asn Ile Asn Ser	
65 70 75 80	
aat gga agt ata tgc ctt gac att ctc aaa gag cag tgg agc cct gct	288
Asn Gly Ser Ile Cys Leu Asp Ile Leu Lys Glu Gln Trp Ser Pro Ala	
85 90 95	
ctg aca att tct aag gtc ctg ctc tcc atc tgc tcc ctg ctg acc gac	336
Leu Thr Ile Ser Lys Val Leu Leu Ser Ile Cys Ser Leu Leu Thr Asp	
100 105 110	
ccc aac ccg gac gac cct ctt gtc ccg gag att gcc cac atg tac aag	384
Pro Asn Pro Asp Asp Pro Leu Val Pro Glu Ile Ala His Met Tyr Lys	
115 120 125	
acg gac cgg ccg aag tac gag tcg acg gcc cgc agc tgg acg cag aag	432
Thr Asp Arg Pro Lys Tyr Glu Ser Thr Ala Arg Ser Trp Thr Gln Lys	
130 135 140	

tac gcg atg g gctga 447
 Tyr Ala Met
 145

<210> 2
 <211> 148
 <212> PRT
 <213> Zea mays

<400> 2
 Met Ala Ser Lys Arg Ile Leu Lys Glu Leu Lys Asp Leu Gln Lys Asp
 1 5 10 15
 Pro Pro Thr Ser Cys Ser Ala Gly Pro Ala Gly Glu Asp Met Phe His
 20 25 30
 Trp Gln Ala Thr Ile Met Gly Pro Pro Asp Ser Pro Tyr Ala Gly Gly
 35 40 45
 Val Phe Leu Val Asn Ile His Phe Pro Pro Asp Tyr Pro Phe Lys Pro
 50 55 60
 Pro Lys Val Ser Phe Lys Thr Lys Val Phe His Pro Asn Ile Asn Ser
 65 70 75 80
 Asn Gly Ser Ile Cys Leu Asp Ile Leu Lys Glu Gln Trp Ser Pro Ala
 85 90 95
 Leu Thr Ile Ser Lys Val Leu Leu Ser Ile Cys Ser Leu Leu Thr Asp
 100 105 110
 Pro Asn Pro Asp Asp Pro Leu Val Pro Glu Ile Ala His Met Tyr Lys
 115 120 125
 Thr Asp Arg Pro Lys Tyr Glu Ser Thr Ala Arg Ser Trp Thr Gln Lys
 130 135 140
 Tyr Ala Met Gly
 145

<210> 3
 <211> 447
 <212> DNA
 <213> Zea mays

<220>
 <221> CDS
 <222> (1)...(442)

<400> 3
 atg gca tcc aaa cgt atc ctc aag gaa tta aag gac ttg caa aag gac 48
 Met Ala Ser Lys Arg Ile Leu Lys Glu Leu Lys Asp Leu Gln Lys Asp
 1 5 10 15
 cct cca aca tca tgc agt gca ggt cct gct ggt gag gat atg ttc cac 96
 Pro Pro Thr Ser Cys Ser Ala Gly Pro Ala Gly Glu Asp Met Phe His
 20 25 30
 tgg cag gca acc atc atg ggt cct cct gat agt cca tat gct gga ggt 144
 Trp Gln Ala Thr Ile Met Gly Pro Pro Asp Ser Pro Tyr Ala Gly Gly
 35 40 45
 gtt ttc ttg gtg aac att cat ttc ccg cca gac tac ccc ttc aag cct 192
 Val Phe Leu Val Asn Ile His Phe Pro Pro Asp Tyr Pro Phe Lys Pro
 50 55 60
 cca aag gta tct ttt aag acg aag gtc ttc cat cca aac atc aat agt 240
 Pro Lys Val Ser Phe Lys Thr Lys Val Phe His Pro Asn Ile Asn Ser
 65 70 75 80

aat gga agc ata tgt ctt gac att ctt aag gag cag tgg agc cct gct 288
 Asn Gly Ser Ile Cys Leu Asp Ile Leu Lys Glu Gln Trp Ser Pro Ala
 85 90 95

 ttg aca atc tct aag gtt ttg ctt tct atc tgc tcc ctg ctt act gat 336
 Leu Thr Ile Ser Lys Val Leu Leu Ser Ile Cys Ser Leu Leu Thr Asp
 100 105 110

 ccc aac cca gac gac cct ctt gtc cct gag att gct cac atg tac aag 384
 Pro Asn Pro Asp Asp Pro Leu Val Pro Glu Ile Ala His Met Tyr Lys
 115 120 125

 acg gat cgg cct aag tac gag gcg aca gcc cgc agc tgg acc cag aag 432
 Thr Asp Arg Pro Lys Tyr Glu Ala Thr Ala Arg Ser Trp Thr Gln Lys
 130 135 140

 tac gcc atg g gatga 447
 Tyr Ala Met
 145

<210> 4
 <211> 148
 <212> PRT
 <213> Zea mays

<400> 4
 Met Ala Ser Lys Arg Ile Leu Lys Glu Leu Lys Asp Leu Gln Lys Asp
 1 5 10 15
 Pro Pro Thr Ser Cys Ser Ala Gly Pro Ala Gly Glu Asp Met Phe His
 20 25 30
 Trp Gln Ala Thr Ile Met Gly Pro Pro Asp Ser Pro Tyr Ala Gly Gly
 35 40 45
 Val Phe Leu Val Asn Ile His Phe Pro Pro Asp Tyr Pro Phe Lys Pro
 50 55 60
 Pro Lys Val Ser Phe Lys Thr Lys Val Phe His Pro Asn Ile Asn Ser
 65 70 75 80
 Asn Gly Ser Ile Cys Leu Asp Ile Leu Lys Glu Gln Trp Ser Pro Ala
 85 90 95
 Leu Thr Ile Ser Lys Val Leu Leu Ser Ile Cys Ser Leu Leu Thr Asp
 100 105 110
 Pro Asn Pro Asp Asp Pro Leu Val Pro Glu Ile Ala His Met Tyr Lys
 115 120 125
 Thr Asp Arg Pro Lys Tyr Glu Ala Thr Ala Arg Ser Trp Thr Gln Lys
 130 135 140
 Tyr Ala Met Gly
 145

<210> 5
 <211> 447
 <212> DNA
 <213> Zea mays

<220>
 <221> CDS
 <222> (1)...(442)

<400> 5
 atg gcg tcg aag cgg atc ctc aag gag ctc aag gac ctg cag aag gat 48
 Met Ala Ser Lys Arg Ile Leu Lys Glu Leu Lys Asp Leu Gln Lys Asp
 1 5 10 15

cct ccc acc tct tgc agc gca ggc cct gtt gcc gaa gat atg ttc tac 96
 Pro Pro Thr Ser Cys Ser Ala Gly Pro Val Ala Glu Asp Met Phe Tyr
 20 25 30

tgg cag gcg acg att atg ggg cca tca gat agc cca tac gct ggt ggc 144
 Trp Gln Ala Thr Ile Met Gly Pro Ser Asp Ser Pro Tyr Ala Gly Gly
 35 40 45

gta ttt ttg gtc act att cac ttt cca ccg gac tac cca ttc aaa cca 192
 Val Phe Leu Val Thr Ile His Phe Pro Pro Asp Tyr Pro Phe Lys Pro
 50 55 60

ccg aag gtt gca ttc aag acg aag gtt tac cat ccg aat atc aac agc 240
 Pro Lys Val Ala Phe Lys Thr Lys Val Tyr His Pro Asn Ile Asn Ser
 65 70 75 80

aac ggg agc atc tgt ctt gat atc ttg aag gag caa tgg agc cct gca 288
 Asn Gly Ser Ile Cys Leu Asp Ile Leu Lys Glu Gln Trp Ser Pro Ala
 85 90 95

ctg aca gtt tct aag gtc ctc ctc tca att tgt tcc ctt ctg acg gac 336
 Leu Thr Val Ser Lys Val Leu Leu Ser Ile Cys Ser Leu Leu Thr Asp
 100 105 110

cca aac cct gac gat ccg ttg gtc cca gag att gct cac ttg tac aag 384
 Pro Asn Pro Asp Asp Pro Leu Val Pro Glu Ile Ala His Leu Tyr Lys
 115 120 125

act gat cgc gtg aag tac gaa tcg act gct agg agc tgg act cag aaa 432
 Thr Asp Arg Val Lys Tyr Glu Ser Thr Ala Arg Ser Trp Thr Gln Lys
 130 135 140

tat gca atg g gctga 447
 Tyr Ala Met
 145

<210> 6
 <211> 148
 <212> PRT
 <213> Zea mays

<400> 6
 Met Ala Ser Lys Arg Ile Leu Lys Glu Leu Lys Asp Leu Gln Lys Asp
 1 5 10 15
 Pro Pro Thr Ser Cys Ser Ala Gly Pro Val Ala Glu Asp Met Phe Tyr
 20 25 30
 Trp Gln Ala Thr Ile Met Gly Pro Ser Asp Ser Pro Tyr Ala Gly Gly
 35 40 45
 Val Phe Leu Val Thr Ile His Phe Pro Pro Asp Tyr Pro Phe Lys Pro
 50 55 60
 Pro Lys Val Ala Phe Lys Thr Lys Val Tyr His Pro Asn Ile Asn Ser
 65 70 75 80
 Asn Gly Ser Ile Cys Leu Asp Ile Leu Lys Glu Gln Trp Ser Pro Ala
 85 90 95
 Leu Thr Val Ser Lys Val Leu Leu Ser Ile Cys Ser Leu Leu Thr Asp
 100 105 110
 Pro Asn Pro Asp Asp Pro Leu Val Pro Glu Ile Ala His Leu Tyr Lys
 115 120 125
 Thr Asp Arg Val Lys Tyr Glu Ser Thr Ala Arg Ser Trp Thr Gln Lys
 130 135 140
 Tyr Ala Met Gly

145

<210> 7
 <211> 459
 <212> DNA
 <213> Zea mays

<220>
 <221> CDS
 <222> (1)...(454)

<400> 7
 atg tcg acg ccg gcg agg aag cgc ctg atg cgg gac ttc aag cgc ctg 48
 Met Ser Thr Pro Ala Arg Lys Arg Leu Met Arg Asp Phe Lys Arg Leu
 1 5 10 15
 cag cag gac ccg ccc gcc ggc atc agc ggc gcg ccg tac gac aac aat 96
 Gln Gln Asp Pro Pro Ala Gly Ile Ser Gly Ala Pro Tyr Asp Asn Asn
 20 25 30
 atc atg ctc tgg aac gcc gtc ata ttc ggg ccg gat gac acg ccg tgg 144
 Ile Met Leu Trp Asn Ala Val Ile Phe Gly Pro Asp Asp Thr Pro Trp
 35 40 45
 gat gga ggc acg ttc aag ctg act ctg caa tgt aca gaa gat tac ccc 192
 Asp Gly Gly Thr Phe Lys Leu Thr Leu Gln Cys Thr Glu Asp Tyr Pro
 50 55 60
 aac aaa cct cca act gtt cgg ttt gtt tcg agg atg ttt cat cca aac 240
 Asn Lys Pro Pro Thr Val Arg Phe Val Ser Arg Met Phe His Pro Asn
 65 70 75 80
 atc tat gcg gat gga agc atc tgt ttg gac atc cta caa aac cag tgg 288
 Ile Tyr Ala Asp Gly Ser Ile Cys Leu Asp Ile Leu Gln Asn Gln Trp
 85 90 95
 agc cct ata tat gat gtt gcg gca ata ttg acc tct att cag tcc ctg 336
 Ser Pro Ile Tyr Asp Val Ala Ala Ile Leu Thr Ser Ile Gln Ser Leu
 100 105 110
 ctg tgt gac cca aac cca aat tct cct gct aac tct gaa gca gct aga 384
 Leu Cys Asp Pro Asn Pro Asn Ser Pro Ala Asn Ser Glu Ala Ala Arg
 115 120 125
 atg ttc agc gag aac aag cgt gag tac aac cgc aaa gtt cgt gag gtt 432
 Met Phe Ser Glu Asn Lys Arg Glu Tyr Asn Arg Lys Val Arg Glu Val
 130 135 140
 gtg gag cag agc tgg aca gcc g actga 459
 Val Glu Gln Ser Trp Thr Ala
 145 150

<210> 8
 <211> 152
 <212> PRT
 <213> Zea mays

<400> 8
 Met Ser Thr Pro Ala Arg Lys Arg Leu Met Arg Asp Phe Lys Arg Leu
 1 5 10 15
 Gln Gln Asp Pro Pro Ala Gly Ile Ser Gly Ala Pro Tyr Asp Asn Asn

	20		25		30
Ile Met Leu Trp Asn Ala Val	Ile Phe Gly Pro Asp Asp Thr Pro Trp				
35	40	45			
Asp Gly Gly Thr Phe Lys Leu Thr Leu Gln Cys Thr Glu Asp Tyr Pro					
50	55	60			
Asn Lys Pro Pro Thr Val Arg Phe Val Ser Arg Met Phe His Pro Asn					
65	70	75			80
Ile Tyr Ala Asp Gly Ser Ile Cys Leu Asp Ile Leu Gln Asn Gln Trp					
85	90	95			
Ser Pro Ile Tyr Asp Val Ala Ala Ile Leu Thr Ser Ile Gln Ser Leu					
100	105	110			
Leu Cys Asp Pro Asn Pro Asn Ser Pro Ala Asn Ser Glu Ala Ala Arg					
115	120	125			
Met Phe Ser Glu Asn Lys Arg Glu Tyr Asn Arg Lys Val Arg Glu Val					
130	135	140			
Val Glu Gln Ser Trp Thr Ala Asp					
145	150				

<210> 9
 <211> 459
 <212> DNA
 <213> Zea mays

<220>
 <221> CDS
 <222> (1)... (454)

<400> 9

atg tcg acg ccg gcg agg aag cgc ctg atg cgg gac ttc aag cgc ctg	48
Met Ser Thr Pro Ala Arg Lys Arg Leu Met Arg Asp Phe Lys Arg Leu	
1 5 10 15	

cag cag gac ccg ccc gcc ggc atc agc ggc gcg ccg tac gac aac aat	96
Gln Gln Asp Pro Pro Ala Gly Ile Ser Gly Ala Pro Tyr Asp Asn Asn	
20 25 30	

atc atg ctc tgg aac gcc gtc ata ttc ggg ccg gat gac acg ccg tgg	144
Ile Met Leu Trp Asn Ala Val Ile Phe Gly Pro Asp Asp Thr Pro Trp	
35 40 45	

gat gga ggc acg ttc aag ctg act ctg caa tgt aca gaa gat tac ccc	192
Asp Gly Gly Thr Phe Lys Thr Leu Gln Cys Thr Glu Asp Tyr Pro	
50 55 60	

aac aaa cct cca act gtt cgg ttt gtt tcg agg atg ttt cat cca aac	240
Asn Lys Pro Pro Thr Val Arg Phe Val Ser Arg Met Phe His Pro Asn	
65 70 75 80	

atc tat gcg gat gga agc atc tgt ttg gac atc cta caa aac aag tgg	288
Ile Tyr Ala Asp Gly Ser Ile Cys Leu Asp Ile Leu Gln Asn Lys Trp	
85 90 95	

agc cct ata tat gat gtt gcg gca ata ttg acc tct att cag tcc ctg	336
Ser Pro Ile Tyr Asp Val Ala Ala Ile Leu Thr Ser Ile Gln Ser Leu	
100 105 110	

ctg tgt gac cca aac cca aat tct cct gct aac tct gaa gca ggt aga	384
Leu Cys Asp Pro Asn Pro Asn Ser Pro Ala Asn Ser Glu Ala Gly Arg	
115 120 125	

atg ttc agc gag aac aag cgt gag tac aac cgc aaa gtt cgt gag gtt	432
Met Phe Ser Glu Asn Lys Arg Glu Tyr Asn Arg Lys Val Arg Glu Val	

130 135 140 459
 gtg gag cag agc tgg aca gcc g actga
 Val Glu Gln Ser Trp Thr Ala
 145 150

<210> 10
 <211> 152
 <212> PRT
 <213> Zea mays

<400> 10
 Met Ser Thr Pro Ala Arg Lys Arg Leu Met Arg Asp Phe Lys Arg Leu
 1 5 10 15
 Gln Gln Asp Pro Pro Ala Gly Ile Ser Gly Ala Pro Tyr Asp Asn Asn
 20 25 30
 Ile Met Leu Trp Asn Ala Val Ile Phe Gly Pro Asp Asp Thr Pro Trp
 35 40 45
 Asp Gly Gly Thr Phe Lys Leu Thr Leu Gln Cys Thr Glu Asp Tyr Pro
 50 55 60
 Asn Lys Pro Pro Thr Val Arg Phe Val Ser Arg Met Phe His Pro Asn
 65 70 75 80
 Ile Tyr Ala Asp Gly Ser Ile Cys Leu Asp Ile Leu Gln Asn Lys Trp
 85 90 95
 Ser Pro Ile Tyr Asp Val Ala Ala Ile Leu Thr Ser Ile Gln Ser Leu
 100 105 110
 Leu Cys Asp Pro Asn Pro Asn Ser Pro Ala Asn Ser Glu Ala Gly Arg
 115 120 125
 Met Phe Ser Glu Asn Lys Arg Glu Tyr Asn Arg Lys Val Arg Glu Val
 130 135 140
 Val Glu Gln Ser Trp Thr Ala Asp
 145 150

<210> 11
 <211> 459
 <212> DNA
 <213> Zea mays

<220>
 <221> CDS
 <222> (1)... (454)

<400> 11
 atg tcg act cct gcg agg aag agg ctg atg agg gat ttc aag cgg ttg 48
 Met Ser Thr Pro Ala Arg Lys Arg Leu Met Arg Asp Phe Lys Arg Leu
 1 5 10 15
 atg cag gac cct ccg gcc ggc ata agt ggc gcc ccg cag gat aac aac 96
 Met Gln Asp Pro Pro Ala Gly Ile Ser Gly Ala Pro Gln Asp Asn Asn
 20 25 30
 ata atg ctg tgg aat gct gtc ata ttt gga cct gat gat acc ccg tgg 144
 Ile Met Leu Trp Asn Ala Val Ile Phe Gly Pro Asp Asp Thr Pro Trp
 35 40 45
 gat gga ggt acg ttt aag ctg act ctc cag ttt aat gaa gaa tat cca 192
 Asp Gly Gly Thr Phe Lys Leu Thr Leu Gln Phe Asn Glu Glu Tyr Pro
 50 55 60
 aac aaa cca cca act gtg cga ttt gtt tct agg atg ttc cat cct aac 240
 Asn Lys Pro Pro Thr Val Arg Phe Val Ser Arg Met Phe His Pro Asn

65	70	75	80	
att tat gct gat gga agc ata tgc tta gat atc ctg cag gat cag tgg				288
Ile Tyr Ala Asp Gly Ser Ile Cys Leu Asp Ile Leu Gln Asp Gln Trp				
	85	90	95	
agt ccg ata tat gat gta gat gcg ata ctc acg tca atc cag tca ttg				336
Ser Pro Ile Tyr Asp Val Asp Ala Ile Leu Thr Ser Ile Gln Ser Leu				
	100	105	110	
ttg tgc gat cca aac ccg aat tcg ccc gct aac tct gaa gcc gcc cgc				384
Leu Cys Asp Pro Asn Pro Asn Ser Pro Ala Asn Ser Glu Ala Ala Arg				
	115	120	125	
atg ttc agc gag aac aag cga gag tac aac cgc aaa gtt cgg gaa gtc				432
Met Phe Ser Glu Asn Lys Arg Glu Tyr Asn Arg Lys Val Arg Glu Val				
	130	135	140	
gtg gag cag agc tgg acg gcg g actga				459
Val Glu Gln Ser Trp Thr Ala				
	145	150		

<210> 12
 <211> 152
 <212> PRT
 <213> Zea mays

<400> 12

1	5	10	15
Met Ser Thr Pro Ala Arg Lys Arg Leu Met Arg Asp Phe Lys Arg Leu			
Met Gln Asp Pro Pro Ala Gly Ile Ser Gly Ala Pro Gln Asp Asn Asn			
	20	25	30
Ile Met Leu Trp Asn Ala Val Ile Phe Gly Pro Asp Asp Thr Pro Trp			
	35	40	45
Asp Gly Gly Thr Phe Lys Leu Thr Leu Gln Phe Asn Glu Glu Tyr Pro			
	50	55	60
Asn Lys Pro Pro Thr Val Arg Phe Val Ser Arg Met Phe His Pro Asn			
	65	70	75
Ile Tyr Ala Asp Gly Ser Ile Cys Leu Asp Ile Leu Gln Asp Gln Trp			
	85	90	95
Ser Pro Ile Tyr Asp Val Asp Ala Ile Leu Thr Ser Ile Gln Ser Leu			
	100	105	110
Leu Cys Asp Pro Asn Pro Asn Ser Pro Ala Asn Ser Glu Ala Ala Arg			
	115	120	125
Met Phe Ser Glu Asn Lys Arg Glu Tyr Asn Arg Lys Val Arg Glu Val			
	130	135	140
Val Glu Gln Ser Trp Thr Ala Asp			
	145	150	

<210> 13
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Designed oligonucleotide based upon an adaptor
 used for cDNA library construction and poly(dT) to
 remove clones which have a poly(A) tail but no
 cDNA insert.

<400> 13
 tcgacccacg cgtccgaaaa aaaaaaaaaa aaaaaa
 36

<210> 14
 <211> 172
 <212> PRT
 <213> Saccharomyces cerevisiae

<400> 14
 Met Ser Thr Pro Ala Arg Arg Arg Leu Met Arg Asp Phe Lys Arg Met
 1 5 10 15
 Lys Glu Asp Ala Pro Pro Gly Val Ser Ala Ser Pro Leu Pro Asp Asn
 20 25 30
 Val Met Val Trp Asn Ala Met Ile Ile Gly Pro Ala Asp Thr Pro Tyr
 35 40 45
 Glu Asp Gly Thr Phe Arg Leu Leu Leu Glu Phe Asp Glu Glu Tyr Pro
 50 55 60
 Asn Lys Pro Pro His Val Lys Phe Leu Ser Glu Met Phe His Pro Asn
 65 70 75 80
 Val Tyr Ala Asn Gly Glu Ile Cys Leu Asp Ile Leu Gln Asn Arg Trp
 85 90 95
 Thr Pro Thr Tyr Asp Val Ala Ser Ile Leu Thr Ser Ile Gln Ser Leu
 100 105 110
 Phe Asn Asp Pro Asn Pro Ala Ser Pro Ala Asn Val Glu Ala Ala Thr
 115 120 125
 Leu Phe Lys Asp His Lys Ser Gln Tyr Val Lys Arg Val Lys Glu Thr
 130 135 140
 Val Glu Lys Ser Trp Glu Asp Asp Met Asp Asp Met Asp Asp Asp Asp
 145 150 155 160
 Asp Asp Asp Asp Asp Asp Asp Asp Glu Ala Asp
 165 170

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/18128

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/82 C12N15/29 C12N9/00 C12N5/10 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YAMAMOTO, K. AND SASAKI, T.: "rice cDNA from callus (970724.1081)" EMBL SEQUENCE DATA LIBRARY, 6 August 1997 (1997-08-06), XP002125331 heidelberg, germany accession no.C29069	1
X	GIROD, P.A., ET AL.: "homologs of the essential ubiquitin conjugating enzymes UBC1,4, and 5 in yeast are encoded by a multigene family in Arabidopsis thaliana" THE PLANT JOURNAL, vol. 3, no. 4, April 1993 (1993-04), pages 545-552, XP002125332 the whole document	11

-/-

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"A" document member of the same patent family

Date of the actual completion of the international search

17 December 1999

Date of mailing of the international search report

11/01/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentplan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer

Holtorf, S

INTERNATIONAL SEARCH REPORT

Int'l. Application No
PCT/US 99/18128

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MYERS, S. AND BAYSDORFER, C.: "untitled" EMBL SEQUENCE DATA LIBRARY, 11 December 1997 (1997-12-11), XP002125333 heidelberg, germany accession no. AF034946	1-11
A	& MYERS, S. AND BAYSDORFER, C.: "untitled" EMBL SEQUENCE DATA LIBRARY - TREMBLREL., 1 June 1998 (1998-06-01), heidelberg, germany accession no. 048555	1-11
A	SULLIVAN, M.L. AND VIERSTRA, R.D.: "cloning of a 16-kDa ubiquitin carrier protein from wheat and Arabidopsis thaliana. Identification of functional domains by in vitro mutagenesis" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 266, no. 35, December 1991 (1991-12), pages 23878-23885, XP002125334 the whole document	1-11
A	ZWIRN, P., ET AL.: "Arabidopsis thaliana RAD6 homolog AtUBC2 complements UV sensitivity, but not N-end rule degradation deficiency, of Saccharomyces cerevisiae rad6 mutants" CURRENT GENETICS, vol. 32, no. 5, November 1997 (1997-11), pages 309-314, XP000863092 the whole document	1-11
A	GENSCHIK, P., ET AL.: "differential expression of several E2-type ubiquitin carrier protein genes at different developmental stages in Arabidopsis thaliana and Nicotiana tabacum" MOLECULAR AND GENERAL GENETICS, vol. 244, no. 5, September 1994 (1994-09), pages 548-556, XP002125335 the whole document	1-11
A	SULLIVAN, M.L., AND VIERSTRA, R.D.: "a ubiquitin carrier protein from wheat germ is structurally and functionally similar to the yeast DNA repair enzyme encoded by RAD6" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 86, no. 24, December 1989 (1989-12), pages 9861-9865, XP002125336 cited in the application the whole document	1-11
	-/-	

INTERNATIONAL SEARCH REPORT

Int'l Application No
PCT/US 99/18128

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 626 450 A (WISCONSIN ALUMNI RES FOUND) 30 November 1994 (1994-11-30) abstract, pages 2,4;examples; claims	1-11
A	WO 94 21794 A (ZENECA LTD ;ABU BAKAR UMI KALSOM (MY); BARTON SARAH LOUISE (GB); G) 29 September 1994 (1994-09-29) abstract, pages 2,7,30-33; Fig. 10; SEQID17; examples 5-12; Table 1; claims	1-11
A	WO 96 16182 A (PIONEER HI BRED INT) 30 May 1996 (1996-05-30) abstract, page 3,24; claims	1-11
A	VARSHAVSKY A: "The ubiquitin system" TIBS TRENDS IN BIOCHEMICAL SCIENCES,EN,ELSEVIER PUBLICATION, CAMBRIDGE, vol. 22, no. 10, page 383-387 XP004091998 ISSN: 0968-0004 figure 1	1-11
P,X	WALBOT,V.: "maize ESTs from various cDNA libraries sequenced at Stanford University" EMBL SEQUENCE DATA LIBRARY, 9 June 1999 (1999-06-09), XP002125337 heidelberg, germany accession no. AI711626	1
P,X	WALBOT, V.: "maize ESTs from various cDNA libraries sequenced at Stanford University" EMBL SEQUENCE DATA LIBRARY, 25 May 1999 (1999-05-25), XP002125338 heidelberg, germany accession no. AI677284	1
P,X	SHOEMAKER, R., ET AL. : "public soybean EST project" EMBL SEQUENCE DATA LIBRARY, 16 March 1999 (1999-03-16), XP002125339 heidelberg, germany accession no. AI461087	1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 99/18128

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this International application, as follows:

See additional sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/ US 99 /18128

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims 1-11 partially.

Nucleotide sequence encoding a maize-specific RAD6-homolog characterized by SEQIDs 3 and 4, respectively. Furthermore, The recombinant expression of the same and a method of modulating the level of endogenous RAD6 activity in a plant by generating transgenic plants expressing the same.

2. Claims 1-11 partially.

Nucleotide sequence encoding a maize-specific RAD6-homolog characterized by SEQIDs 5 and 6, respectively. Furthermore, The recombinant expression of the same and a method of modulating the level of endogenous RAD6 activity in a plant by generating transgenic plants expressing the same.

3. Claims 1-11 partially.

Nucleotide sequence encoding a maize-specific RAD6-homolog characterized by SEQIDs 7 and 8, respectively. Furthermore, The recombinant expression of the same and a method of modulating the level of endogenous RAD6 activity in a plant by generating transgenic plants expressing the same.

4. Claims 1-11 partially.

Nucleotide sequence encoding a maize-specific RAD6-homolog characterized by SEQIDs 9 and 10, respectively. Furthermore, The recombinant expression of the same and a method of modulating the level of endogenous RAD6 activity in a plant by generating transgenic plants expressing the same.

5. Claims 1-11 partially.

Nucleotide sequence encoding a maize-specific RAD6-homolog characterized by SEQIDs 11 and 12, respectively. Furthermore, The recombinant expression of the same and a method of modulating the level of endogenous RAD6 activity in a plant by generating transgenic plants expressing the same.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Application No

PCT/US 99/18128

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 0626450	A	30-11-1994	JP 7147987 A	13-06-1995
			US 5851791 A	22-12-1998
WO 9421794	A	29-09-1994	AU 687793 B	05-03-1998
			AU 6262394 A	11-10-1994
			CA 2158473 A	29-09-1994
			EP 0689594 A	03-01-1996
			JP 8507923 T	27-08-1996
			US 5908973 A	01-06-1999
WO 9616182	A	30-05-1996	US 5633438 A	27-05-1997
			AU 701202 B	21-01-1999
			AU 4194496 A	17-06-1996
			CA 2205505 A	30-05-1996
			EP 0797674 A	01-10-1997
			HU 78060 A	28-07-1999
			NZ 297145 A	25-11-1998
			US 5756324 A	26-05-1998

THIS PAGE BLANK (USPTO)